

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using SW model

Run on: March 22, 2006, 22:57:33 ; Search time 56 Seconds
(without alignments)
2291.668 Million cell updates/sec

Title: US-10-736-769-4

Perfect score: 6909
Sequence: 1 MAEAGIRGWLMLLRLAQLA.....GSIKAGALSINPLNNRQPF 1332

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 445344 seqs, 96347055 residues

Total number of hits satisfying chosen parameters: 445344

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

- 1: Pending_Patents_AA_New.*
- 2: /cgn2_6/prodata/1/paa/PCT_NEW_COMB.pcp.*
- 3: /cgn2_6/prodata/1/paa/US06_NEW_COMB.pcp.*
- 4: /cgn2_6/prodata/1/paa/US07_NEW_COMB.pcp.*
- 5: /cgn2_6/prodata/1/paa/US08_NEW_COMB.pcp.*
- 6: /cgn2_6/prodata/1/paa/US09_NEW_COMB.pcp.*
- 7: /cgn2_6/prodata/1/paa/US10_NEW_COMB.pcp.*
- 8: /cgn2_6/prodata/1/paa/US60_NEW_COMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6909	100.0	1332	6 US-11-301-094-6	Sequence 6, Appl1
2	6872.5	99.5	1359	6 US-11-270-796-22	Sequence 22, Appl1
3	5407	78.3	1333	6 US-11-270-796-3	Sequence 3, Appl1
4	2402.5	34.8	1278	1 PCT-US06-05584-822	Sequence 822, App
5	2402.5	34.8	1278	6 US-11-191-274A-129	Sequence 129, App
6	2402.5	34.8	1278	6 US-11-191-274A-130	Sequence 130, App
7	1341.5	19.4	1383	6 US-11-301-094-2	Sequence 2, Appl1
8	1046	15.1	1274	6 US-11-301-094-4	Sequence 4, Appl1
9	782.5	11.3	419	6 US-10-953-449-21065	Sequence 21065, A
10	588.5	8.5	967	6 US-10-461-673-16739	Sequence 16739, A
11	583.5	8.4	1182	6 US-11-332-764-2	Sequence 2, Appl1
12	583.5	8.4	1182	6 US-11-337-244-149	Sequence 149, App
13	482	7.0	891	8 US-60-772-265-1197	Sequence 1197, App
14	455	6.6	1358	8 US-60-772-265-241	Sequence 241, App
15	437	6.3	831	7 US-11-360-355-133805	Sequence 133805, A
16	349	5.1	783	6 US-10-461-673-16750	Sequence 16750, A
17	325	4.7	204	7 US-11-360-355-141351	Sequence 141351, A
18	261	3.8	575	7 US-11-360-355-120785	Sequence 120785, A
19	252	3.6	505	6 US-10-461-673-16871	Sequence 16871, A
20	229.5	3.3	465	7 US-11-360-355-141367	Sequence 141367, A
21	225	3.3	542	6 US-11-214-063A-1670	Sequence 1670, App
22	222.5	3.2	422	7 US-11-360-355-141365	Sequence 141365, A
23	219	3.2	1137	6 US-10-461-673-16889	Sequence 16889, A
24	219	3.2	1330	6 US-10-461-673-10541	Sequence 10541, A
25	190	2.8	126	7 US-11-360-355-141352	Sequence 141352, A

26	152.5	2.2	632	6 US-11-293-697-3548	Sequence 3548, App
27	141	2.0	164	7 US-11-360-355-127069	Sequence 127069, A
28	139.5	2.0	484	7 US-11-360-355-120169	Sequence 120169, A
29	135.5	2.0	252	7 US-11-360-355-140206	Sequence 140206, A
30	123.5	1.8	171	7 US-11-360-355-149589	Sequence 149589, A
31	121.5	1.8	1043	6 US-10-536-606-20	Sequence 20, Appl1
32	120.5	1.7	2426	6 US-11-203-806A-11	Sequence 11, Appl1
33	118.5	1.7	768	6 US-11-214-063A-2044	Sequence 2044, App
34	117.5	1.7	788	6 US-11-214-063A-1692	Sequence 1692, App
35	117.5	1.7	619	8 US-60-732-162-1828	Sequence 1828, App
36	116.5	1.7	619	6 US-11-312-958-46	Sequence 46, Appl1
37	116.5	1.7	619	6 US-10-461-673-12167	Sequence 12167, A
38	114	1.7	488	8 US-60-752-353-45206	Sequence 45206, A
39	114	1.7	985	6 US-11-293-697-2874	Sequence 2874, App
40	112.5	1.6	697	6 US-10-703-799B-226	Sequence 226, App
41	111	1.6	201	7 US-11-045-004-2027	Sequence 152528, A
42	110.5	1.6	494	6 US-11-360-355-152528	Sequence 2027, App
43	109.5	1.6	451	6 US-11-045-004-2398	Sequence 2398, App
44	109.5	1.6	3979	8 US-60-752-353-33716	Sequence 33716, A
45	108	1.6	474	8 US-60-752-353-39891	Sequence 39891, A

ALIGNMENTS

RESULT 1
US-11-301-094-6
; Sequence 6, Application US/11301094
; GENERAL INFORMATION:
; APPLICANT: Levitan, Diane J
; TITLE OF INVENTION: FUNCTIONAL ASSAYS FOR CHOLESTEROL ABSORPTION INHIBITORS
; FILE REFERENCE: JBO6242US01
; CURRENT APPLICATION NUMBER: US/11/301,094
; PRIOR FILING DATE: 2005-12-12
; PRIOR FILING DATE: 2004-12-15
; NUMBER OF SEQ NOS: 24
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 6
; LENGTH: 1332
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-301-094-6

Query Match 100.0%; Score 6909; DB 6; Length 1332;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1332; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MAEAGIRGWLMLLRLAQLAQBEPYTIHQGYCAFYDECGKPELSGSLMTLSNVSLSN	60
DB	1	MAEAGIRGWLMLLRLAQLAQBEPYTIHQGYCAFYDECGKPELSGSLMTLSNVSLSN	60
QY	61	TPARKITGDHLILLOKICPRPLYGENTQACSAKOLVLSLASITKALITRCPACSDNF	120
DB	61	TPARKITGDHLILLOKICPRPLYGENTQACSAKOLVLSLASITKALITRCPACSDNF	120
QY	121	VNIHCNTPSPNOSLFINTRVNOAGOLPAVVAEAFQHSFAROSYDSCSRVPPA	180
DB	121	VNIHCNTPSPNOSLFINTRVNOAGOLPAVVAEAFQHSFAROSYDSCSRVPPA	180
QY	181	ATLAVGTCGVYSALCNARWLNFGDGTNGLAIPDITPHLEPGQAVSGGIQPLNEGV	240
DB	181	ATLAVGTCGVYSALCNARWLNFGDGTNGLAIPDITPHLEPGQAVSGGIQPLNEGV	240
QY	241	ARCNEQGDVATCSCQDCAACPAPALDSTYLYLGMPGSLVLIILGCVFAVVTI	300
DB	241	ARCNEQGDVATCSCQDCAACPAPALDSTYLYLGMPGSLVLIILGCVFAVVTI	300
QY	301	LIVGFVAPARDSKRVDPKKGTSLSFSTHTLLGQFQCGMTWVSWPLTIIVLAVS	360
DB	301	LIVGFVAPARDSKRVDPKKGTSLSFSTHTLLGQFQCGMTWVSWPLTIIVLAVS	360

QY 361 IIVVLAAGLVTELTTPVEELMSAPNSQARSEKAFHDOHFQFFRTNOVILTAENRSSY 420
DB 361 IIVVLAAGLVTELTTPVEELMSAPNSQARSEKAFHDOHFQFFRTNOVILTAENRSSY 420
QY 421 RYDSLLGPKNSGILDLLELELQERLRLHLOWSPBAORNISLODICAPLNPNT 480
DB 421 RYDSLLGPKNSGILDLLELELQERLRLHLOWSPBAORNISLODICAPLNPNT 480
QY 481 SLYDCINSLOYFQNNRTLLLTANQTMGOTSQVDMKDHFLYCANAPLTFKDGATALL 540
DB 481 SLYDCINSLOYFQNNRTLLLTANQTMGOTSQVDMKDHFLYCANAPLTFKDGATALL 540
QY 541 SCMAUYGAVPFPFLAIGYKKGKDYSEAEALIMTFSLNNYPADPRLAOKLMEAELEEM 600
DB 541 SCMAUYGAVPFPFLAIGYKKGKDYSEAEALIMTFSLNNYPADPRLAOKLMEAELEEM 600
QY 601 RAFORMMAGFOVTFPAERSLDEINRTTAEDLPFAISYIVIFLYISLALGSYSMSRV 660
DB 601 RAFORMMAGFOVTFPAERSLDEINRTTAEDLPFAISYIVIFLYISLALGSYSMSRV 660
QY 661 MVDSKATLGLGVAVVLGAVMAAMGFPSYLGIRSSLVILQVVPFLVLSGADNIFIFVLE 720
DB 661 MVDSKATLGLGVAVVLGAVMAAMGFPSYLGIRSSLVILQVVPFLVLSGADNIFIFVLE 720
QY 721 YQRLPRRPGEPREVIHGRALGRVAPSMILCSLSEALCFPLGALTMPAVRTFALTSLGLAV 780
DB 721 YQRLPRRPGEPREVIHGRALGRVAPSMILCSLSEALCFPLGALTMPAVRTFALTSLGLAV 780
QY 781 ILDFLLQMSAFVALLSLDSKROBASRLDVCCCVKPOELPPGQGBGLLGFQKAYAPPL 840
DB 781 ILDFLLQMSAFVALLSLDSKROBASRLDVCCCVKPOELPPGQGBGLLGFQKAYAPPL 840
QY 841 LHMTRGVLLFLALFGVSLYSMCHISVGLDQELALPKDSYLLDYFLFLNRYFEVGA 900
DB 841 LHMTRGVLLFLALFGVSLYSMCHISVGLDQELALPKDSYLLDYFLFLNRYFEVGA 900
QY 901 YVVTLLGYNFSSBAGMNAICSSAGCNSFSTQKIQTATEPPEOSYIATPASSVDDPFI 960
DB 901 YVVTLLGYNFSSBAGMNAICSSAGCNSFSTQKIQTATEPPEOSYIATPASSVDDPFI 960
QY 961 LPPSSCCRLYISGPNKDKFCPTVNSLNCCKNMSITMGSVRSVQFHKYLPWFINDRP 1020
DB 961 LPPSSCCRLYISGPNKDKFCPTVNSLNCCKNMSITMGSVRSVQFHKYLPWFINDRP 1020
QY 1021 NIKCPKGLAAYSTSVNLTSDGVLASRFMAVHKPKNSQDYTEALRAARELANITADL 1080
DB 1021 NIKCPKGLAAYSTSVNLTSDGVLASRFMAVHKPKNSQDYTEALRAARELANITADL 1080
QY 1081 RYVPGTDPAFEVPPYITTVFEBQYITLPEGLFMLSCLVPTFAVSCULLGLDLSGILL 1140
DB 1081 RYVPGTDPAFEVPPYITTVFEBQYITLPEGLFMLSCLVPTFAVSCULLGLDLSGILL 1140
QY 1141 NLLSIWMLVDYTGFMALWDISYNAVSLINLVSAGMSVEFVSHITRSRISTKPTWLER 1200
DB 1141 NLLSIWMLVDYTGFMALWDISYNAVSLINLVSAGMSVEFVSHITRSRISTKPTWLER 1200
QY 1201 AKBATISMSGSAFVAGVAMTNLPGIIVLGLAKAQLIOIFFEFLNLTLLTLLGLHGLFLEV 1260
DB 1201 AKBATISMSGSAFVAGVAMTNLPGIIVLGLAKAQLIOIFFEFLNLTLLTLLGLHGLFLEV 1260
QY 1261 IISYGVDPVNPALALEQKAEBAVAAVMVASCNHPRSVSTADNIYVNSHFBGSIKAGA 1320
DB 1261 IISYGVDPVNPALALEQKAEBAVAAVMVASCNHPRSVSTADNIYVNSHFBGSIKAGA 1320
QY 1321 ISNPLPNNGROF 1332
DB 1321 ISNPLPNNGROF 1332

RESULT 2
US-11-270-796-22
; Sequence 22, Application US/11270796
; GENERAL INFORMATION:

; APPLICANT: Dong, Jianli
; TITLE OF INVENTION: RESTORATION OF CANCER BY SIMULTANEOUS INHIBITION OF BRAF AND
; FILE REFERENCE: 02420/1201581-US1
; CURRENT APPLICATION NUMBER: US/11/270,796
; PRIOR FILING DATE: 2005-11-08
; PRIOR APPLICATION NUMBER: 60/592,592
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 22
; LENGTH: 1359
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-270-796-22

Query Match 99.5%; Score 6872.5; DB 6; Length 1359;
Best Local Similarity 97.9%; Pred No. 0;
Matches 1330; Conservative 0; Mismatches 2; Indels 27; Gaps 1;

QY 1 MAEAGLRGMLWALLRLAQBEPYTTIHQPGYCAEYDEGKNPBLSGIMTLSNVSCLSN 60
DB 1 MAEAGLRGMLWALLRLAQBEPYTTIHQPGYCAEYDEGKNPBLSGIMTLSNVSCLSN 60
QY 61 TPARKITGDHLLILQKICPRLYTGPNTOACCSAKQULYSIABSLSTKALTTCPCACSDNF 120
DB 61 TPARKITGDHLLILQKICPRLYTGPNTOACCSAKQULYSIABSLSTKALTTCPCACSDNF 120
QY 121 VNLHCNHTCSPNQSLFINVTRVAQAGQLPVAVYEAFOHSFABOSYDSCSRVRVPA 180
DB 121 VNLHCNHTCSPNQSLFINVTRVAQAGQLPVAVYEAFOHSFABOSYDSCSRVRVPA 180
QY 181 ATLAVGTMCGVYSALCNAORMLNFQDGTGNGLABLDITFHLBEGQAVSGSIQPLNEGV 240
DB 181 ATLAVGTMCGVYSALCNAORMLNFQDGTGNGLABLDITFHLBEGQAVSGSIQPLNEGV 240
QY 241 ANCNESQGDVATCSCODCASCPAIAAPQALDSTFYLGOMGSLVLIILCSVAVVTI 300
DB 241 ANCNESQGDVATCSCODCASCPAIAAPQALDSTFYLGOMGSLVLIILCSVAVVTI 300
QY 301 LTVGRVVPAPARKSKMVPBKGTSLSDKLSFSTHTLLGQFQGMGNWASWPLTILVLSV 360
DB 301 LTVGRVVPAPARKSKMVPBKGTSLSDKLSFSTHTLLGQFQGMGNWASWPLTILVLSV 360
QY 361 IIVVLAAGLVTELTTPVEELMSAPNSQARSEKAFHDOHFQFFRTNOVILTAENRSSY 420
DB 361 IIVVLAAGLVTELTTPVEELMSAPNSQARSEKAFHDOHFQFFRTNOVILTAENRSSY 420
QY 421 RYDSLLGPKNSGILDLLELELQERLRLHLOWSPBAORNISLODICAPLNPNT 480
DB 421 RYDSLLGPKNSGILDLLELELQERLRLHLOWSPBAORNISLODICAPLNPNT 480
QY 481 SLYDCINSLOYFQNNRTLLLTANQTMGOTSQVDMKDHFLYCANAPLTFKDGATALL 540
DB 481 SLYDCINSLOYFQNNRTLLLTANQTMGOTSQVDMKDHFLYCANAPLTFKDGATALL 540
QY 541 SCMAUYGAVPFPFLAIGYKKGKDYSEAEALIMTFSLNNYPADPRLAOKLMEAELEEM 600
DB 541 SCMAUYGAVPFPFLAIGYKKGKDYSEAEALIMTFSLNNYPADPRLAOKLMEAELEEM 600
QY 601 RAFORMMAGFOVTFPAERSLDEINRTTAEDLPFAISYIVIFLYISLALGSYSMSRV 660
DB 601 RAFORMMAGFOVTFPAERSLDEINRTTAEDLPFAISYIVIFLYISLALGSYSMSRV 660
QY 661 MVDSKATLGLGVAVVLGAVMAAMGFPSYLGIRSSLVILQVVPFLVLSGADNIFIFVLE 720
DB 661 MVDSKATLGLGVAVVLGAVMAAMGFPSYLGIRSSLVILQVVPFLVLSGADNIFIFVLE 720
QY 721 YQRLPRRPGEPREVIHGRALGRVAPSMILCSLSEALCFPLGALTMPAVRTFALTSLGLAV 780
DB 721 YQRLPRRPGEPREVIHGRALGRVAPSMILCSLSEALCFPLGALTMPAVRTFALTSLGLAV 780
QY 781 ILDFLLQMSAFVALLSLDSKROBASRLDVCCCVKPOELPPGQGBGLLGFQKAYAPPL 840

[illegible]

Db	123	IHCNHTCSBDQSLEFNTRVYQRPBGQLEPVAVAEAFQNSFEAKENSCSRVNI PAAS	182
Qy	183	IAVGTMCCVYSALCNAQRMNLFOGDTNGLAPLDITPHLEPGQAVSGGIQPIINEGVAR	242
Db	183	LAVSGMCGVYSALCNAQRMNLFOGDTNGLAPLDITPHLEPGQALADGKKPLDGKITP	242
Qy	243	CNCSGQDDVATCSCODCASCPRAPARPOLDSTFYLGMQPSLVYIILCSFVAVTILL	302
Db	243	CNCSQGBEBSAACSODCASCPRVIPPFPALRPSFWMGMPMMLIIIFTVAVILLSVL	302
Qy	303	VGFRVAPARBDKXWMDPKKGSLSBDKXSFSEHTLLIGQFQMGMGTVASWPLTIVLSVIP	362
Db	303	VYLRASNRKNKKTAGSOEAPNLPKRRFSFHTVLGRFESWGTRVASMPLTVLALSTIV	362
Qy	363	VVALAGVLFTELTTDPVELMSAPNSQARSEKAFHDOHFGFPFRFTQVYLTAPNRSSRY	422
Db	363	VIALSVGLTFLBELTTDPVELMSAPNSQARKKKAFFHDFHGFPPFTNQFVYAKORSSYKY	422
Qy	423	DSLILGPNKFSGIIDLLELLELLELOERLRLQWSPBAQNSISLODI CYAPLNDNTSL	482
Db	423	DSLILGPNKFSGIISLDLQELLELOERLRLQWSEHAQNSISLODI CYAPLNPHTSL	482
Qy	483	YDCCNSILQYFONNRITLLLTANOITLGMQNSOVDWMDHFLYCANAPLTFPDGALALSC	542
Db	483	TDCCNSNSLQYFONNHTLLLTANOITLNGQTSLDVMDKHFLYCANAPLTYDGTALALSC	542
Qy	543	MADYCAPPFPFATIGYKGDYSEAEALIMFSLNNYPAGDPRLAQALMEAPFLEEMRA	602
Db	543	IADYCAPPFPFLAVGTYGCTDYSERALKITFSSINNPADDPKRAHAKLMEAPFLKEMQS	602
Qy	603	FORMAGMFOYTFTAERSLEDEINRTTAEDLPITATSYIVIFLYISLGSYSMSRWVY	662
Db	603	FQSRSTADRFQJAFSAERSLEDEINRTTQIDLPVFASIVLYIFLYISLGSYSRMSRVAV	662
Qy	663	DSKATLIGGVAVVLGAMMAAGPFSYIGISSLVILQVFPFLVSGANDIIFTVLEQY	722
Db	663	DSKATLIGDGVAVVLGAVVAAAGFYSYIGVSSLVIIQVFPFLVAVADANI FTFVLEQY	722
Qy	723	RLPRRGEPREVIHGRALGRVAPSMVLCSTSEALCFPLGALTMPPAVTFPLTSGIAVIL	782
Db	723	RLPRRGEPQREAHIGRTIGSVAPSMVLCSTSEALCFPLGALTMPPAVTFPLTSGIAIIF	782
Qy	783	DFLLQMSAFVALLSIDSKROBASRLDYCCVCKPOELPPQGBEGLLGFQKAYAPFLH	842
Db	783	DFLLQMTAFVALLSIDSKROBASRPDYCCSSRNLPKQKEGILLCFKXIYTFPFLH	842
Qy	843	WITRGVILLFLAFGVSLYSMCHI SVGLDOELAPKOSYLLDYFLPLNRYFEVGAARYF	902
Db	843	RFIRFVVLLEFLVFGANVLYMCNISVGLDODLAPKOSYLLDYFLPLNRLYEVGPARYF	902
Qy	903	VYTLGYNPSSERAGMNAICSSAGCNNSFTOKIOYATEPEBOSYALPASSVNDPIMLT	962
Db	903	DTTSYNSNSTEAGMNAICSSAGCSSFSLTQIKIYQASEPBNOSYVYALASSVNDPIMLT	962
Qy	963	P-SSCCRLYIISGPNKDFCPSFTVNSLNLCKXCMSITMGSVYSPSEOFHXYLPWFLENDPN	1021
Db	963	PSSCCRLYITGPHHDEFCEPSTDBNSFNCLCKXCMRTIGPVAPFTBQHFXYLPWFLENDPN	1022
Qy	1022	IKCPRGGLAAYSTSVNLTSDSQVTLASRFMAVHKPLKNSDYTEALRAAREBLAANIYTDLR	1081
Db	1023	IRCPRGGLAAYRTSVNLSSDQIITASQPMAYHKPLRNSQDTEALRASRLAANIYTAELR	1082
Qy	1082	KVPGDPAFVFPYPIYTNVFPYBOYLTILBGLFMLSILCVPTFAVSCLLGLDIBRSGLTN	1141
Db	1083	KVPGDPAFVFPYPIYSNVFQOYLYTVLBEGIFPTALCEVTFPVCYLLGLDIBRSGLTN	1142
Qy	1142	LLSIYMIIVDTVGFEMALWDISYNAVSLINTVSAVGMSEVFSHTITRSFAISTKPTWLERA	1201
Db	1143	LLSIIMILVDTIIGLAAVGISYNAVSLINTVYANVGMSEVFSHTITRSFAVSTKPTBLERA	1202
Qy	1202	KEATISMGSAVPAGVAMTNLPGLVGLAKAQIIFPFRNLNLTLLGLHGLVFLPYI	1261

Db 1203 KDATFMGSAVAGVAMTNPGLILIGFAQAOLIOIFFERRLMLITLLGLHGVLPV 1262
Qy 1262 ISYVGDVNPALALEQKRAEEVAAMVASCNHPERSVSTADNTYNSFEES-IXGAGA 1320
Db 1263 ISYLGPDVQALVLEEKLAATEA-AMVSEPSCPQYPPADANTSVDYNYGFNDEFIPEINA 1321
Qy 1321 ISNFIENNGROF 1332
Db 1322 ASSSLPKSDOKP 1333

RESULT 4

PCT-US06-05584-822
; Sequence 822, Application PC/TUS0605584
; GENERAL INFORMATION:
; APPLICANT: DANA-FARBER CANCER INSTITUTE, INC.
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION,
; FILE OF INVENTION: ASSESSMENT, PREVENTION, AND THERAPY OF CANCER
; FILE REFERENCE: Dfs-064.25(12592-6425)
; CURRENT APPLICATION NUMBER: PCT/US06/05584
; CURRENT FILING DATE: 2006-03-02
; PRIOR APPLICATION NUMBER: 60/690,064
; PRIOR FILING DATE: 2005-06-13
; PRIOR APPLICATION NUMBER: 60/654,227
; PRIOR FILING DATE: 2005-02-17
; NUMBER OF SEQ ID NOS: 848
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 822
; LENGTH: 1278
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US06-05584-822

Query Match 34.8%; Score 2402.5; DB 1; Length 1278;
Best Local Similarity 39.8%; Pred. No. 1.1e-185;
Matches 522; Conservative 227; Mismatches 450; Indels 103; Gaps 25;

Qy 7 RGMILMALILRLAQSEPTTTHQPCYCARVDECG-----KNPELSSGLMLTSLNVSCLIS 59
Db 4 RGLALGLLLLLCPAQVFSQ-----SCWYGGCGIAYGDKRYNCEYSG----- 46
Qy 60 NTPARKITGDHLLLOKICPRLYTGPNTQACCSAKVLSEASLSTTKALTRCPACSDN 119
Db 47 --PYKLPDQYDLVDELCPGFPG--NGLCCDVRLQTLKONLQPLQPLSLRCSGCFPN 103
Qy 120 FVNLCHNTCSBNOSLFINVTR-----VAQLGAGQLPAAVAAFYQHSFAEOSYDSCSRV 175
Db 104 LNLFCGLTCSBRSQSLNVTATEDVVDVTNQTNNKVELQYVVGQSFAANMAYNACRPV 163
Qy 176 RVPAAATLAVGTMCVGYGALCNAQRWLNFGDGTGNGLAPLDT-----PHLEPGQA 228
Db 164 EAPSSNDKALGLCGKDADA-CNATWIEYMFKNKGQAPFTITPVFSDPVH----- 215
Qy 229 VSGGIQPLNEGVARCNESGDDVATGSCODCAASCALARPQ-----ALDSTFVLG 279
Db 216 ---GMEPMNATKGCDESDVETAPCSCODCSIVGCPKQPPPPAPMTLLGLDAMYVIM 272
Qy 280 QMPGSLVILILICSYFAVVTILL-----VGRVAPARDKSKMVPKKGSTLS 326
Db 273 WITWAFLLVFGAFPAVWCYKRKYVSYTPIDSIANSV-NASKGE-----ASGC 324
Qy 327 DKLSTSTHTLLGQFQSGKGTWASWPLTLVLVSIVVALAAGLVTELTTPDVELMASP 386
Db 325 DVSVAFAFEGCLRLRFRMGSCFCVRNCGVIFFSLVFITACSSGLVFRVTTNPVDLMSAP 384
Qy 387 NSQARSEKAFHOHQGFPRTRNOVILITANRSRYRDLGLPK-NFSGIILDLDEL 445
Db 385 SSQARLEKXYFOHGFPRTRTEQLIRALPTDKHLYQYPSGADVFGPPLDIQLIHQVL 444
Qy 446 ELQERLRHLQVWSPEAQRNLSLQDICYAPLNPDTSLYDCCINSLLQYFQNNRTLLTLTA 505
Db 445 DLQAIEN--ITASYNETVTIQLDICALPLSPYNT--NCTILSVANTYQNSHSLVDHKK 499

Qy 506 NOTLMGQTSQVMDKHFLYCANAPLTFKDGTAALASCMADGAPVFPFLAIGYKGDYS 565
Db 500 GDDF---VYADYHTHTFLYCVRAPASLNDTSLHDPCLGTGCGVFPMLVLGGYDDQNTN 556
Qy 566 EAEALIMTFSLNINYPADGPRLAQAKLMEAEFLMEARAFQRMAGMFQVTFTAERSLDEI 625
Db 557 NATALVITFPVNNYNDTEKIQRAQAKEKEFINPKYKN---PULTISFTAKRSIBDL 613
Qy 626 NRTTAEADLPITATSYIVFLYISLALSGSYSSKSRVMSKATLGLGVAUVLGAAMAAG 685
Db 614 NRESDVFTVVISYAIMFLYISLALGHIKSCRLLVDSKSLGJAGILYLVSSVACSIG 673
Qy 666 FFSYGISSIVILLOVPEVLVSAGDNIFIIVLEYORLPRRGRBREHIGRALGRVAP 745
Db 674 VFSYIGLPTLIVIPPLVAVGVNDIFILVQAYORDERQGETLQOQLGRVLGEVAP 733
Qy 746 SMLTSLSEALCFPLGALTTPPAVATFALTSGLAVIDLFLQMSAFVALSLDSKROBAS 805
Db 734 SMFLSFSBEYAFIAGALSNVPAVHTFSLPAGLANFIDFLQITCFVGLLGLDIRQEN 793
Qy 806 RLIVCCCVKPOELPPPGQ-GEGLLGFQKAYAPFLHITRGVLLFLALFGVLSYSM 864
Db 794 RLDFCCVRGAEDGTSVQASESCLEFRFKNSYSPILLKDMRPVIAIFVGLSFSIAVL 853
Qy 865 CHISVGLDQELALPKDSYLLDYFLFLNRYFVGAVYFVTLGNVFSSEAGNNALCSSAG 924
Db 854 NKVDIGLQDSJMPDQSYWVDYFKSISQYLHAGPVPYVLEBGHDTYSKQGNMVCGGNG 913
Qy 925 CNMFSTOKIQYATEFPQSYLAIIPASSWVDYFIDMLTP-SGCCRLYISGPNKDFCPSY 983
Db 914 CNNDLSLVQOIFPAQALNDYTRIIGFAPSIMIDYFPWVAPQSSCCV---DNITDQFNAS 970
Qy 984 VNSLNLCKNCSIT-MGSVRPSVEQFKYLLPMLFNDRNKICPKGGLAISTSVN--LTS 1040
Db 971 VVDPAQCV-CBPLTFEBGQRQCGDFMRFLPMLFSDNPNPCGKGHAAYSANVILIGH 1029
Qy 1041 DQVLAASFMAVHRKLNKSDYTEALRAARELANITADLRKVPETDAFEVFPYTTNV 1100
Db 1030 GTRVATYFMVHTVLTQTSADPTDLKRAKRLASVNT-ETNGINS--AYRVFYSYV 1086
Qy 1101 FVEQYLTLLPEGLFMLSCLVPTFAVSCLLGLDRLSGLLNLDSIVMLLVDTVGFMALMD 1160
Db 1087 FVEQYLTIIIDTIFLVLGSLAIFLVTMVLGCELSAVINCATIAMVLNMFQWMLJG 1146
Qy 1161 ISYNAVSILINISAVGMSVEFVSHITRSFAISTKPTWLERAKERTISGSAVAVAMTN 1220
Db 1147 ISLNAVSILINISVMSGSISEFCSHITRAFTVMSKMSRVERAEALAHMGSSVFSGITLTK 1206
Qy 1221 LPLGILVGLAKAQLQIFFRRLNLLITLLGLHGVLPVILSYGPPVNA 1272
Db 1207 FGGIVVLAFAKSQIFQIFFRVYILAMVLLGATHGLIFLPVLLSYGPSVNA 1258

RESULT 5

US-11-191-274A-129
; Sequence 129, Application US/11191274A
; GENERAL INFORMATION:
; APPLICANT: Appleera Corporation
; APPLICANT: Bruno DOMO
; TITLE OF INVENTION: LUNG DISEASE TARGETS AND USES THEREOF
; FILE REFERENCE: C1001536PROV
; CURRENT APPLICATION NUMBER: US/11/191,274A
; NUMBER OF SEQ ID NOS: 334
; SOFTWARE: FaastSeq for Windows Version 4.0
; SEQ ID NO 129
; LENGTH: 1278
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-191-274A-129
Query Match 34.8%; Score 2402.5; DB 6; Length 1278;
Best Local Similarity 39.8%; Pred. No. 1.1e-185;

Matches 522; Conservative 237; Mismatches 450; Indels 103; Gaps 25;

QY 7 RGMWLLALLRLAQSERYTTHIQGYCAFYDECG-----KNPELSGSLMTLSNVSCLS 59
 Db 4 RGLALGILLLLCPAIVFSQ-----SCWVGEGCIAYGDKRYNCEYSG----- 46

QY 60 NTPARKITGDHLLLOKICPRLYTGPTTQACSAKQVLSLEASITTKALLTRCPACSDN 119
 Db 47 --PPKPLPKDGYDLVBLCEGPFPG--NVSLLCCVROLQTKMLQPLQPLSRCPSCFYN 103

QY 120 FVNHLCHNTCSPNQSLFINVTR-----VAOLGAGQLPAAVAYEAFOHSFAEGSYDSCSRV 175
 Db 104 LNLFLCELTCSPRQSFPLNTATEDYVDPVTNQTNNVKELOYYVGQSFANAMYNACRDV 163

QY 176 RVPAAATLAVGTWCGYYSALCNAORWLNFGQDTGNGLAFLDIT-----FHLBFGQA 228
 Db 164 EAPSSNDKALGLLCCGDADA--CNATWIEFMKDGQAPFTITTPVSDPVA----- 215

QY 229 VSGGIQPLNEGVARCNESQGDVATCSQODCAASCPAIAAPQ-----ALDSTFYLG 279
 Db 216 ---GMEPPNNATKGCDESDVEVTAAPCSQDCSIVCCPKPQPPPPAPWTLIGLDAMYIM 272

QY 280 QMRGSLVLIILICSVAVVTILL-----VGRVAPARDKSKMVDKKGTSLS 326
 Db 273 WTTWAAFLVFGAFPAVWCYRRRYVSEYTPIDSNIAFSV--NASDKG-----ASCC 324

QY 327 DKLSFSTHTLLGCFQFGMGTWASWPLTILVSVIPVVALAAGLVFTELTDDPELMSAP 386
 Db 325 DVPASAFEGCLRLFTRWGSCFCVRNPGCVIFPSLVITACSSGLVFRVYTNVDLMSAP 384

QY 387 NSQARSEKAFHDQHFPPFTNOVILTAENRSSYRYSLLGPK--NFSGILDDLLELL 445
 Db 385 SSQARLEKEYFDHFGPPFTBOLIRAPLTDKHIQYPPSGADVFPGLDQILHQVL 444

QY 446 ELQERLRHLQWSPBAORNISLQDICAPLNPNTSLYDCINSLLQYFQNNRTLLLLTA 505
 Db 445 DLQIALEN--ITASYNETVTLQDICIAPLSPYNT--NCTILSVLNFYSHNSVADHKK 499

QY 506 NQTLMOQTSQVDKDFLYCANAPLTFKQGTALALSCMADYGAVPFPLAIGCYKGDYS 565
 Db 500 GDDF---YVADYHTFLCYVRAPASLNTSLHDCELTGFGGPFVFWLVLGGYDDQNTN 566

QY 566 EABALIMTSLANNYPAGDPRLAQAKLMEAFLEMBRPFORRMAGMFQVTFARSLDEI 625
 Db 557 NATALVTFEPVNNYVNDTEKLGQAQWKEKFINFVKYKN--PNLTISFTARSJEDL 613

QY 626 NRTTADLPIFATSYVIFYISLAGSYSSMRWVDSKATIGLGVAVVLCAMWAMG 685
 Db 614 NRESDDVFTVVISYAIMPLDYLISLAGHKSRRLLVDSKVSIGIAGILVLSSVACSLG 673

QY 686 FFSYLGIRSSLVLYOVVPLVLSVGADNPIFVLEYQRLPRRGEPREVHIGRALRVAP 745
 Db 674 VSYISLPLTLIYEVIFPLVLAAGVNDIFILVQAYRDERLQGETLDOQLGVLGEVAP 733

QY 746 SMLLCSLSEALCFELGATPMPAVRTFALTSGVLVLDLQMSAFVALSLDSKQEAS 805
 Db 734 SWEFSFSEVNAFELGLASMPAVHTFSLFAGLAVFIDLQITCFVSLGLDILIKQEKX 793

QY 806 RLDVCCCVKPOELPPGQ--GEGILLGFQKAYAPFLMLHTTRGVLLFLALFGVLSYM 864
 Db 794 RLDIFCCVGAEGDTSVQASECLPFPFKNSYSPLLKMMRPDIVAIFGVLSFSIAVL 853

QY 865 CHISVGLDDELALPKOSYLLDLYFLPLNRYPEVGAPVYFVTTLGYNFSSEAGMMAICSSAG 924
 Db 854 NKVDIDGLDLSLMPDSSIMVDYFKSISQYIHAAPRYFVLEBHDITSSGQMMVCGMG 913

QY 925 CNFSTOKIQVATEEPQSYLAIPASSWVDFIDWLTLP--SSCCRLYISGPNKDKCPST 983
 Db 914 CNNDLSLVQOIFNAAQDNTYTRIGFAPSSWIDYFDWVKPQSSCCRV---DNITDQCNAS 970

QY 984 VNSLNLKMKMSIT--MGSVRPSVEQHFHXYLPFLNDRPNKICPKGGLAAYSTSVN--LTS 1040
 Db 971 VVDPACVR--CRPLTPGKQRPQGGDFMRFLPMFLSDNPNKCGKGGHAAVSSAVNILLGH 1029

QY 1041 DGOVLASRFMAHYKPLKNSODYTEALPAARELANITADLRKYPGTDPAEVEPYTITNV 1100
 Db 1030 GTRVGATYEMTHTVLTQTSADPDLAKKARLISANT--ETMGINGS--AYRVPPYSFVY 1086

QY 1101 FYEBOYLTILPEGLFMTSLCVPFPAVSCILLGIDLSGLINLSIWMILVDTVGPMLMD 1160
 Db 1087 FYEBOYLTIDDTTFNLGVSIGALFVYTWLLGCELSAVIMCATIAVAVLVNMGVMMKMG 1146

QY 1161 ISYNAVSLINLVASAVGSEVFNHITRSFALSTKPTMLERAKBATISMSGAVAGVAMTN 1220
 Db 1147 ISLNAVSLVNLWMSGSIYVEFCHITRAFVSKSGRVEABEALAHMSSSVSGITLTK 1206

QY 1221 LPGILVGLAKAQOLIQIFFERLNLITLLGLHLGVFLPVILSYGPDVNA 1272
 Db 1207 FGGIVVLAFAKSQIFQIFYFRMYLAVLLGATGILFLPLVLSYIGSVAKA 1258

RESULT 6
 US-11-191-274A-130
 ; Sequence 130, Application US/11191274A
 ; GENERAL INFORMATION:
 ; APPLICANT: Applera Corporation
 ; APPLICANT: Bruno DOMO
 ; TITLE OF INVENTION: LUNG DISEASE TARGETS AND USES THEREOF
 ; FILE REFERENCE: CL001536PROV
 ; CURRENT APPLICATION NUMBER: US/11/191,274A
 ; CURRENT FILING DATE: 2005-07-28
 ; NUMBER OF SEQ ID NOS: 334
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 130
 ; LENGTH: 1278
 ; TYPE: PR1
 ; ORGANISM: Homo sapiens
 US-11-191-274A-130

Query Match 34.8%; Score 2402.5; DB 6; Length 1278;
 Best local Similarity 39.8%; Pred. No. 1,1e-185;
 Matches 522; Conservative 237; Mismatches 450; Indels 103; Gaps 25;

QY 7 RGMWLLALLRLAQSERYTTHIQGYCAFYDECG-----KNPELSGSLMTLSNVSCLS 59
 Db 4 RGLALGILLLLCPAIVFSQ-----SCWVGEGCIAYGDKRYNCEYSG----- 46

QY 60 NTPARKITGDHLLLOKICPRLYTGPTTQACSAKQVLSLEASITTKALLTRCPACSDN 119
 Db 47 --PPKPLPKDGYDLVBLCEGPFPG--NVSLLCCVROLQTKMLQPLQPLSRCPSCFYN 103

QY 120 FVNHLCHNTCSPNQSLFINVTR-----VAOLGAGQLPAAVAYEAFOHSFAEGSYDSCSRV 175
 Db 104 LNLFLCELTCSPRQSFPLNTATEDYVDPVTNQTNNVKELOYYVGQSFANAMYNACRDV 163

QY 176 RVPAAATLAVGTWCGYYSALCNAORWLNFGQDTGNGLAFLDIT-----FHLBFGQA 228
 Db 164 EAPSSNDKALGLLCCGDADA--CNATWIEFMKDGQAPFTITTPVSDPVA----- 215

QY 229 VSGGIQPLNEGVARCNESQGDVATCSQODCAASCPAIAAPQ-----ALDSTFYLG 279
 Db 216 ---GMEPPNNATKGCDESDVEVTAAPCSQDCSIVCCPKPQPPPPAPWTLIGLDAMYIM 272

QY 280 QMRGSLVLIILICSVAVVTILL-----VGRVAPARDKSKMVDKKGTSLS 326
 Db 273 WTTWAAFLVFGAFPAVWCYRRRYVSEYTPIDSNIAFSV--NASDKG-----ASCC 324

QY 327 DKLSFSTHTLLGCFQFGMGTWASWPLTILVSVIPVVALAAGLVFTELTDDPELMSAP 386
 Db 325 DVPASAFEGCLRLFTRWGSCFCVRNPGCVIFPSLVITACSSGLVFRVYTNVDLMSAP 384

QY 387 NSQARSEKAFHDQHFPPFTNOVILTAENRSSYRYSLLGPK--NFSGILDDLLELL 445
 Db 385 SSQARLEKEYFDHFGPPFTBOLIRAPLTDKHIQYPPSGADVFPGLDQILHQVL 444

QY 446 ELQERLRHLQWSPBAORNISLQDICAPLNPNTSLYDCINSLLQYFQNNRTLLLLTA 505

Db 974 SPECYVYVDPN--TFGSTNRKNSALDDKACRTCMDFDYVANSYPKSSIMYHRPSIEVEY 1031
Qy 1010 KTLPMFLNDRPNKICPKGGLAAYSTSVNLTSQGVLASRFMAVHKRL--KNSODYTEAR 1067
Db 1032 RHLRHLLETPTNSECEVFGGRASPKDISFTSRGRLOASQPMTHFKLSTSSNSDFTKAD 1091
Qy 1068 AARELANITADLRKVPGTDPAPFEVPTITNVFEQVYLTLLPEGLFMLSCLVFPFAVS 1127
Db 1092 TBAWTSRRLERSI-----DOTAHVAVAKSIFPFYQYSTIMPLTQLQPIYVGVFGII 1145
Qy 1128 CULLGLDRSGLLNLSIYMLVDVGFPMALMDISYNAVSLNLVSAGVMSVEFVSHIR 1187
Db 1146 CTTGLIDVGAACAVICQVSNYFHYAFMYIENIPYNALSATNLVMSGILIEFSVNLK 1205
Qy 1188 SFAITKTPWLRKAERATISMSGSAVAGVAMTLPGLVLGLAKAQLQIFPRLNLLT 1247
Db 1206 GYACSLRORAKRABESTVSGIPITILSGVVYVWAGSTMFLSGAHLOITVYFPKPLITI 1265
Qy 1248 LGLLHGLVFLPVLISYVGPDPVNPALALBQKRAEAAVAAVMVASCPNHSRSTADNIVY 1307
Db 1266 VSSAVHALIITLILAFGSGRHSSESTINDNEQHDACVLS--PTASHSISNVEGIL 1323
Qy 1308 N-----HSFEGSIKGAGAI 1321
Db 1324 NRPLLDASHIIDLPLKAGGI 1345
RESULT 8
US-11-301-094-4
Sequence 4, Application US/11301094
APPLICANT: Levitan, Diane J
APPLICANT: Smith, Marsha
TITLE OF INVENTION: FUNCTIONAL ASSAYS FOR CHOLESTEROL ABSORPTION INHIBITORS
FILE REFERENCE: JB062420S01
CURRENT APPLICATION NUMBER: US/11/301, 094
CURRENT FILING DATE: 2005-12-12
PRIOR APPLICATION NUMBER: 60/636,390
PRIOR FILING DATE: 2004-12-15
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn version 3.3
SEQ ID NO 4
LENGTH: 1274
TYPE: PR1
ORGANISM: Caenorhabditis elegans
US-11-301-094-4
Query Match 15.1%; Score 1046; DB 6; Length 1274;
Best Local Similarity 25.7%; Pred. No. 1,2e-75;
Matches 345; Conservative 241; Mismatches 587; Indels 168; Gaps 46;
Qy 10 LMALLLRAGSEPYTTIHQPGYCAFYDGGKVPGLSGIMTLNSVCLSNTPARKITGD 69
Db 14 VFLLLIHLALCO-----AKCVN--TECDGSEDSNHPCKTKNXTYLPITVTRSLNPT 64
Qy 70 HLILQICPRPYTGPN--TOACCSAKQVLSLEASLITRALLTRCPACSDNPFNLHCNT 128
Db 65 WNAPEKYSYLVQEBDKQVCTELQKGMMDRISNAATILGSCSCFDFNFKKMQCPT 124
Qy 129 CSPNOSLFINTRVAOLGAGQLPAVVAEAFYQHSFSAEOSYDSCSRVVPAAATLAVGTM 188
Db 125 CSPDQSKFMKVMET-----TGPKVVVVWMEFKVNRDVEGLYESCRHTWFGANGLRLMSL 180
Qy 189 CGVYSGALCNAQRMLNFOGDTG--NGLA---PLDITHLEPGQAVSGIQPLNEGVARCN 244
Db 181 GGRVVS-----FENFYGMGTKNLAOSIPINTEQFSRMKANMIPTTP-----CH 225
Qy 245 ESQGDVATCSQDCD---AASCPAIRPOLDSTFYLGWPGSLVLIILCSVFVAVTLL 301
Db 226 KAGRPVPAAGALDCTPMAHQVLDISKVHTLGTKVHPHP--DFEWLAKICGCLA-LTYL 283
Qy 302 LVGF-----RVAPARDKSKVNDPKKGTSLSDKLSFSTHTLLGQFFQG---WGTWVAS 350

Db 284 LVFILLKYSCHRBSAPNGEDGCYVDLQGN-----LEVOFEGLCARYANAVIK 330
Qy 351 WPLTLLVLSVIVVALLAAG-LVFTELTTPDVELMSAPNSQASKEAF-HDHFQGPFFRTN 408
Db 331 HPLIFVSLIVAAACCSNPFHSLTHSVDOVASADGSTRREKFKFHS--FGPHRIE 388
Qy 409 QVILTAPNRSSYRYDSLILGPPNFGSILDLDBLLELLEOERLRIHQVMSPEAQRNISIQ 468
Db 389 QIFINLPPT-----KSMFMPLPEENFQLVGNIGNLT--ACYGNSSVKD 432
Qy 469 DICVAPLNDPNTSLDCCINSLLQYFQNNRTLLLTANOTLMGQTSQVDMKHFLYCANA 528
Db 433 DICYKRIQGN-----HGCALMSPTNYFQNKMTFENAGPPTIDEIDEDDQWHELKYCIRN 488
Qy 529 PLTFQDGTALALSCMADYCAPVPPFLATGG-----YKGXQ--YSEBAELIMPTSLNNYPAGD 583
Db 489 PLTV--STYSEMSCFEESFGPIDPLLVFGSNEBSIGAMMYTARTIMTVLRG-----541
Qy 584 PRLAQKLWEEAFLBEMRAFORRMAGMPQVTFEASLEDEINRTTAED--LPIFATSYI 641
Db 542 PE-DQAIAMETAFPLNMSRYEMKHANF---TMTETSVABEIHVAVETDKIVSIVACAAV 597
Qy 642 VIFLYISLALGSY--SSWSRVVDSKATLGLGVAVAVLCAVMAAMGFSLYIGRSSLVIL 699
Db 598 LTMVITMLGINHPRESSILSALVHHKLISISAVMISVISVSCSIGMFSLFGVHARDMAI 657
Qy 700 QVPPFLVLSVGADNIFIFLVEYQRLPRRGEF---REVN--IGRLGVABSMLLCSLS 753
Db 658 VVLFVITCTGIRNFVLIIRTFQANGCYGLPNISYREKMHRI SNWRASIPVLNLSLI 717
Qy 754 EALCPFL-GALTP-----MPAVRTPALTSGLAVILDFLLQMSAFVALLSLSKRO-----802
Db 718 CSTCLFAGGVLPYVSVMPAVEVFRHAGLIMDTAAYLVMLLPQYDAREMSGKC 777
Qy 803 -----EASRLDVCCEKQDELPPPCQGBGLLGFQKAYAPFLMHTTRGVALLP 853
Db 778 EIMPWYELSNESKINKINCEAVDGNLRP-----VDWKALIAPLLLKKICIKWTATPF 830
Qy 854 LALFGVSLYSKCHISVGLDQELALPKDSYLLDFLEFLNRYFEVGADEVYVTTLTGINPSS 913
Db 831 FVSLIACYCTCLBERGFQVNAFSETSYLTGKFQNMENLNIIGPLMFEVEGDDVKHPD 890
Qy 914 AGMNAICSSAGCANNBFTQKIQ---YATFPPOSYLAIRASSVNDPFIQMLTP--SSCCRL 969
Db 891 KMQNKFCITLGGDDSMGNKIRSLAAYENY-KGNYLHGDVNIWLDSDYLOFMHRRGSCCKM 949
Qy 970 YISGPKDKFC--PSTVNSLNCUKNCKSITMGSVRPSVEQFHKYLPFLNDRPNKICPKGG 1028
Db 950 -----DGKQFCDS--NATHC--SSCSSSVASLITTEYEFYRNLHFLTPPSIQCAHGG 1001
Qy 1029 LAAVSTSVNLTSQGYLASRFMAVHKP--LKNODYTEALRAARELANITADLRKVPGT 1086
Db 1002 MALAKAINALTNNGKIQSAVFSFPFKKLNLSQIQLYDMRPAKYLADIDIERL-EIPGV 1060
Qy 1087 DPAFEVPTITNVFEQVYLTLLPEGLFMLSCLVTPFVASCLLGLDRSGLLNLSY 1146
Db 1061 ----KVVYSTFEPYEQVLTISTVYTLVVVLFAFAFTTISFLTEVNLNLAGSLVTFEVL 1116
Qy 1147 MLVDTGVGMALMDISYNAVSLNLVSAGVMSVEFVSHITRSPASTKTPWLRABEART 1206
Db 1117 SSYLIMEMMYVLLGITVNVVSVINMMSLGIAVEFGQMLHGVNKKPGRBRARALAV 1176
Qy 1207 SMGSAVAGVAMTNPGLIV---LGLAKAQLQIFPRLNLLTLLGLLHGLVFPVLI 1262
Db 1177 SMGSTLSGI-----FALMITAGCLSPADSRVLIITFCNQVIGIGVCAVHGVVPTLL 1232
Qy 1263 SYVGPDPVNPALALBQKRAEEA 1283
Db 1233 AIFGSDFYQNVASSESTDEA 1253
RESULT 9
US-10-953-349-21065

```
/ Sequence 21065, Application US/10953349
/ GENERAL INFORMATION:
/ APPLICANT: ALEXANDROV, Nikolai et al.
/ TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
/ FILE REFERENCE: 2750-1579PUS2
/ CURRENT APPLICATION NUMBER: US/10/953,349
/ CURRENT FILING DATE: 2004-09-30
/ NUMBER OF SEQ ID NOS: 40252
/ SOFTWARE: PatentIn version 3.3
/ SEQ ID NO: 21065
/ LENGTH: 419
/ TYPE: PRT
/ ORGANISM: Glycine max
US-10-953-349-21065

Query Match      11.3%; Score 782.5; DB 6; Length 419;
Beet Local Similarity 38.4%; Pred. No. 5.3e-55;
Matches 164; Conservative 88; Mismatches 138; Indels 37; Gaps 10;

QY      863 SMC-HISVGLDQELALPKDSYLLDFLFLNRYFEVGPARYFTTLGYNFSEA-GMNAIC 920
DB      2  ALCTIEAGLEQOIALPRDSYLGQYFSNISEYLKRGDPILYVK-DYNVLESKHTNOLC 60
QY      921 SGAGCNNSFTOKIQYATFPEQSYLAIPASSWVDDFDLWTPS--SCRLYISGPNKDK 978
DB      61  SISHDSDNSLNEISRASLVPSTSYAKPAASWLDLFLWISPEAFSCCRKF---TND5 116
QY      979 PCPSVNSLNCCKNCKMSITMSV-----RPVEOPHKLPMFLNDRPNI 1022
DB      117 YCPDDQPCCLPDEQPCGLGGVCKDCTTCFPHSDLVNDRPSTAQREKLPFLDLPSA 176
QY      1023 KCPKGLAAYSTSVNLT--SDQVLASRFMAVHKPKLNSQDYTEALRAARELANITADL 1080
DB      177 DCAKGHGAYTNSVDLNGEGVIAQSEERTYHTPLNRGDVYNAIRARDSAISSSL 236
QY      1081 RKVGTDPAEVFPYTTINVPYEQYTLIPBGLFMLSCLVPTPAVSCILLGLDLSGLL 1140
DB      237 K-----MDIFPYSVFYIFFEQYLDIKLALINITVALGAIFV-CLITSSVSSSAI 287
QY      1141 NLSTVILVDTVGFALMDISYNAVSLINIVSAVMSSEFVSHITRSPALSTKPTWLER 1200
DB      288 LLLVIMITLIDMGVAILGIQLNNAVSNLISGIAVEFCVHVHAFVAVSLGDR-SGR 346
QY      1201 AKEATISMSAGVAFAGVAMTNLPGLVLGLAKAQLIQIFFRLNLTLLGLLGLVFLPV 1260
DB      347 ACTALCTMGASVFSGTLTKLVGVLVLCSTSEIPVYVYFQWYALVIIGFLHGLVFLPV 406
QY      1261 ILSVGP 1267
DB      407 VLSLFGP 413

RESULT 10
US-10-461-673-16739
/ Sequence 16739, Application US/10461673
/ GENERAL INFORMATION:
/ APPLICANT: Tang, Y. Tom
/ APPLICANT: Liu, Chenghua
/ APPLICANT: Zhou, Ping
/ APPLICANT: Asundi, Vinod
/ APPLICANT: Wang, Jian-Rui
/ APPLICANT: Ren, Feiyan
/ APPLICANT: Zhang, Jie
/ APPLICANT: Zhao, Qing A.
/ APPLICANT: Xue, Aidong
/ APPLICANT: Wehrman, Tom
/ APPLICANT: Wang, Zhi Wei
/ APPLICANT: Yang, Yonshong
/ APPLICANT: Goodrich, Ryle W.
/ APPLICANT: Wang, Dunrui
/ APPLICANT: Yamazaki, Victoria
/ APPLICANT: Ujwal, Manusha L.
```

```
/ APPLICANT: Ma, Yuning
/ APPLICANT: Chen, Rui-Hong
/ APPLICANT: Ghosh, Malabika
/ APPLICANT: Weng, Gezhi
/ APPLICANT: Haley-Vicente, Dana
/ APPLICANT: Drenthac, Radoje T.
/ TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
/ FILE REFERENCE: 823
/ CURRENT APPLICATION NUMBER: US/10/461,673
/ CURRENT FILING DATE: 2003-06-13
/ PRIOR APPLICATION NUMBER: PCT/US02/29964
/ PRIOR FILING DATE: 2002-09-19
/ PRIOR APPLICATION NUMBER: US 10/245,014
/ PRIOR FILING DATE: 2002-09-16
/ PRIOR APPLICATION NUMBER: US 60/323,739
/ PRIOR FILING DATE: 2001-09-19
/ PRIOR APPLICATION NUMBER: PCT/US02/29636
/ PRIOR FILING DATE: 2002-09-18
/ PRIOR APPLICATION NUMBER: US 10/245,817
/ PRIOR FILING DATE: 2002-09-16
/ PRIOR APPLICATION NUMBER: US 60/323,349
/ PRIOR FILING DATE: 2001-09-18
/ PRIOR APPLICATION NUMBER: PCT/US02/29001
/ PRIOR FILING DATE: 2002-09-13
/ PRIOR APPLICATION NUMBER: US 10/243,552
/ PRIOR FILING DATE: 2002-09-12
/ PRIOR APPLICATION NUMBER: US 60/322,511
/ PRIOR FILING DATE: 2001-09-13
/ PRIOR APPLICATION NUMBER: PCT/US02/25485
/ PRIOR FILING DATE: 2002-08-09
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 17116
/ SOFTWARE: pc_fl_genes Version 6.0
/ SEQ ID NO 16739
/ LENGTH: 967
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-461-673-16739

Query Match      8.5%; Score 588.5; DB 6; Length 967;
Beet Local Similarity 21.3%; Pred. No. 1e-38;
Matches 216; Conservative 192; Mismatches 35; Indels 24; Gaps 37;

QY      334 HT-----LLQFPQSGMTVNASWPLTILVSLVVALAAGLVF--TELTTDPVSLWAP 386
DB      115 HTDCLLEGLSRFFQMLGQVGAHPWIFLLAPLMLTALGTGFLYLPKDEEBDEBHYTPV 174
QY      387 NSQARSEKAFHDQHGPRFRNTQVILTPNRSS-----YRDSLLGPKNFSGIL 436
DB      175 GSPAKAERRFVQGH---FTTNDSYRFSASRSTEFANFVSLVVSDDLDPATFAVS 230
QY      437 DLDLLLELELOERLRLQVWSPKARNISLQDIC--YAPLNPDNTSLYDCI--NSLLQ 492
DB      231 KLDGAVQDLRV-----AREKSGIQYQVQCARFAL-----CVPPNPILY 270
QY      493 YFQNNRTLLLTANQTLMGQTSQVDWKDHFLYCANAPLTFKDGITALALSCNADYCA-EVF 551
DB      271 AMQVKTILNL-----SSISFPAY-----NHRHRLY 296
QY      552 PFLAIGY-----KGQDYSEAKLIMTFSLNTPADDPRL-AQAKIMEAFLEEMKAF 603
DB      297 LTGFFGYITLGGSLMGQLLRAKAMRLLYLK--TEDPEYDVQSKMLTHLLDQFTNI 353
QY      604 ORRMA-----GMFOVTFARSLDEINRTAEDLPFATSYIVFLYISL 649
DB      354 KNIILAKTIBVPGVGLGGQGEKVHFTSLRQLEFEFTSVTVFVFLAIVLITLFAVT 413
QY      650 ALGYSYSRWVNSKATLGLGVAIVLGAVMAAMGFPSSYLGIRSLVILQVVPFLVLSV 709
DB      414 SCFRFD-----CIRKMCVAAFGVISAFVAVSGGLLHIGV-DFVITVANSPLILGV 467
QY      710 GADNIFIVLEYQRLPRRPGEPREVHIGBALGVAVPSMLLSALCIFLGAALTPMPAV 769
```

```

Db      468 GVDMMFIMISAMHK--TNLAGDIRE--RMSVVSKAASVITITITITINILALYTMSFSRV 525
Oy      770 RTFALTSLGLAVLIDPLLQMSAFVALISLDSKQOZSRLDVCCVKPQELP----- 819
Db      526 QCFCIYGTGTLTLFCFYFNITCFGAFMALDGKREV-----VLCWTKKADPKPMSFKKFC 580
Oy      820 -----PQCGEGEL--LLGFQKAVAPFLMHITGRVLLFLALFGVSLYMSCHISVL 871
Db      561 FPFQSVDPBEHGDIHPMSLFFRDYFGPLITRBSKTFVFTVLYLTISSLYCFHQBEL 640
Oy      872 DQELALPKDSYLLDYFLFLNRYF-EVGAPVYEVITLGYNFSSEAGNNAICSSAGCNFSP 930
Db      641 DLRLNASDSDSYTPYFNVEBENYFSDYGRVWVITVKVDYMDK-----DV 685
Oy      931 TOKIOYAEFPQOSTIALP--ASSWDDPDIWLITBSCCRLYISGPNKDKCPSTVNSLN 988
Db      686 RQKENCCKRIPEKKNVYVDKNLTFEFLDADYQ-----YIKGNSQD---PNEKNT-- 730
Oy      989 CLKKCMSTMGSVRSVEQFHXYLPMFLNDRNINICKPKGGLAAGSRYVLTSDDGVLASR 1048
Db      731 -----FNANNIDPFLSNFNP-----FQHDNISSNSNEIISR 761
Oy      1049 FMAYHKPLKNSQDYTEALRAARELANITADLRKYPGTDPAFEVPPYITTNVFEYOYLT 1108
Db      762 GPQIOTDVSS-----AKKI---LLFQLRRI-AEDCQPLMWYNGAFYFPQYAL 809
Oy      1109 LPE-----GLFMLSCLVP-----TFVASCILLGLDLRSGLANLISVILV 1150
Db      810 LEDTVRNVLVASAMFIVSLDIPYPLCSLWTFAGSYIVY----- 852
Oy      1151 DTGFMALMDISYNAVSLINLVSAGMSVEFVSHITRSPAISTKPTMLERAKETISMGS 1210
Db      853 --TGMATMKVNLDSISMLNVITCIGFSDFSAHISYAVASSQSSVAKSVEALYLG 910
Oy      1211 AVFAGVAMTNLEGLIYVLGLAKAQLIQIEFFRNLITLTLGLHGLVLEVLISYVG 1266
Db      911 PVLQGS-AISTIGVCVLAAKAKAYIFRT--PFKIMFLVMIGAHLGIFIVFLTPFG 964

RESULT 11
US-11-332-764-2
; Sequence 2, Application US/11332764
; GENERAL INFORMATION:
; APPLICANT: Wiscotzkey, Robert G.
; TITLE OF INVENTION: PATCHED HOMOLOG 2 (PTCH-2) DISRUPTIONS, COMPOSITIONS AND METHODS
; FILE REFERENCE: R1673 CIP/75658,.064000
; CURRENT APPLICATION NUMBER: US/11/332,764
; CURRENT FILING DATE: 2006-01-13
; PRIOR APPLICATION NUMBER: US 60/413,543
; PRIOR FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: US 10/669,143
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 2
; LENGTH: 1182
; TYPE: PRT
; ORGANISM: Mus musculus
; US-11-332-764-2

```

```

Query Match      8.4%; Score 583.5; DB 6; Length 1182;
Beet Local Similarity 24.2%; Pred. No. 3,6e-38;
Matches 282; Conservative 154; Mismatches 433; Indels 297; Gaps 44;

QY 330 SFSSTTLGQ-----FPGQ---WGTWASMPETLTVSVLPVVALAGLVFTET 376
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 20 SSAPHILGSLGAPLMLRAYFOGLFSLGCRIOKHGKFLGLVAFGALALGRVAVIE 79
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 377 TDPVELMSPNQASSEKAFHDQHG-PPFRINQVILTPNNSSRYKIDLLGKPNFSOI 435
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 80 TDEQLMWEVSGRSQELHYTEKLGEEAAYISOMI-----OTAHDEGNVILTP-----A 131
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

436 LDLDLLELELELQSRHLHQLQWSPDAQRNISLDQICY---APLNPDN-----TSLVDC 486
432 LD-----LHLQALTRASKQVSLYKSKMDLAKTICTYSGVPLLENGHETEMIEKLPVCV 184
487 INSLIQYEQNNRRTLLLLTANQTLMGQTSQVDMKD---HFLYCANAPLT---FXDGTA 539
185 ILTFLDLCWBGAK---LQGSAYLPGRPDIQMTNLDPQQLLEBLQFPASLEGPRELLDKA 241
487 INSLIQYEQNNRRTLLLLTANQTLMGQTSQVDMKD---HFLYCANAPLT---FXDGTA 539
540 LSCNADYGAPV-----PFLA---IGYKGDYS----- 565
242 QVGQAYVGRPCLDDEDPHCPSPAENRHSRQAPNVAOELSGGCHSFHKFMHMOEBLLIG 301
566 -----BAEKLIMTF-----SINNNYAGPRLAQAALMEAFLEEBKARAPORMAG 609
302 TARDLQQLRAEALQSTFELMSFRQYEHHRGQYTHDIGMSEEQASMTVLAQMRKRVQ 361
610 MEQVFTFAERSLE-----DEINRTTAAEDLPITAFSYIV--IFLYISLALGSYSWS 658
362 LAQERLPAHNSQQLHAHSSTTLDDILAFSE---VSTRVVGGLYMLLAACYTMLRMD 417
659 RVMVDSKATLGLGVAVVVLGAVMAMGPFYSIGIRSSLVILQIVQVPLVLSVAGDNIFIV 718
418 --CAQSGAVAGLAVLVALAVASGLGICALLGITFMAATQVPLPFLALGIVDDIFLA 475
719 LEYQRLPRRQCEPRERVHIGRALGVAPASMLCSISEAIPCFLGALTMPMPAVRTALTSGL 778
476 HAFKAP--PPTPLPERHGECLSTRGTSTVALTSVNNVAFPMALVLPILPALRAFSLOAAI 533
779 AVIILDELQMSAPFALLSLDSKQREABRLDYCCC-----VYRQE----- 817
534 VVGNCPAAVMLVFPAIISLDRHRHQRLDVLCCFSSPBCAQVITQMLPQELGDRAVPVGI 593
818 -----LPP-----PG-----QSGG-- 826
594 AHLRTATQAFTHCEASSQHVVTILPQAHLSPASDPLGSLYSRQGSRTROLLQOEBGT 653
827 -----LILGFQKAVAPVFLIMHTRGVLLFLPALFGVSLYSMCHISVGLDQ 873
654 POACRPLCAHMTLHAFARYQFAPLLIQTPAKALVLLFFGALLGLSLYGATLVODDGAL 713
874 ELAPKQSYLLDYFLFLNRYEVBQAPVYFVYTLTGYNSS--SGAGNNAISSAGCNNSFTQ 932
714 TDVPERGKKEHAFISAQLRYFSL--YEVALVYQGGFADHNSQRLA-----FDLHQ 761
933 KIQVATERPEQSYLAIPAS-----SWVDF-----IDMLTSSCCRLTISGPNK 976
762 RFSSLL-----KAVLPPTATQAPKRWTLHYSNLSLOGICQAPRPDPAASGRITHTSRNGSD 816
977 D-----KCPSPSTVNSLNCIKMCSITM-----GSVRPSVEQFAKYLPMFLNDRPNIKCP 1026
817 GALAVKLLIQGMNQEPDLPFSQLTTRKLVDBEGILP--BLFVYGLTVWVSDDL----- 869
1027 GGLAA-----YST---SVNLTSQOVLASRFMAVYHKPLKNSQDYTEALR 1067
870 -GLAASQANFPPEPEMLHDXYDTGEBELRIIPAQPLRFAPQPFLLHGLQCTAFVEAIE 928
1068 AARELANI--TADLRKYPCGIDPAEVEVPYITTNVFEYQOYLILBEGHFMLSIC--LVPTF 1124
929 GARAACTEAGAGVHAAYPSGSP-----LFWQOYGL--RRCCLLAVCLLCLCTF 976
1125 AVSCILLGLDGRSGILNLSIVMLIVDTGVGFMAAMDLSYNVAVSLINVAAGMSVEFEVSH 1184
977 LVCHALLILSPYTAGIT-VLYLAAMTVBELFGIMGLGIKLSAIPVILIVASIGIEVEFTVH 1035
1185 ITRSFALSTKPTWLERAKEATISMGSAVAGV---AMTNLPGLIIVGLAKAQLIQIFFR 1241
1036 VALGFLTSHSGRNIRAA-----SALBQTFAPVTDGAVALSTLGLIMLAGSNPDFIIRFEV 1090
1242 LNLILITLIGLHGLVPLPVILSYGCP 1267
1091 VLVTVTLTGLHGLLPLPVLLSTLGP 1116

```
RESULT 12
US-11-337-244-149
; Sequence 149, Application US/11337244
; GENERAL INFORMATION:
; APPLICANT: William Matthews
; APPLICANT: Mark Moore
; APPLICANT: Russell Phillips
; APPLICANT: Michael V. Wiles
; APPLICANT: Thadd C. Reeder
; APPLICANT: Robert G. Wiscotzkey
; APPLICANT: Keith D. Allen
; APPLICANT: Helen Barbault
; APPLICANT: Thomas J. Brennan
; APPLICANT: Catherine Guenther
; APPLICANT: Robert Klein
; APPLICANT: Christopher J. Kirk
; APPLICANT: Simon X. Xie
; APPLICANT: Qin Zhang
; APPLICANT: Agnes Chopin
; TITLE OF INVENTION: Targeted Gene Disruptions, Compositions
; FILE REFERENCE: R-DB-13
; CURRENT FILING DATE: 2006-01-20
; PRIOR APPLICATION NUMBER: US/11/337,244
; PRIOR FILING DATE: 2003-09-23
; PRIOR APPLICATION NUMBER: US 60/413,647
; PRIOR FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: US 60/413,666
; PRIOR FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: US 60/413,653
; PRIOR FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: US 60/413,646
; PRIOR FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: US 60/413,625
; PRIOR FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: US 60/413,639
; NUMBER OF SEQ ID NOS: 237
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 149
; LENGTH: 1182
; TYPE: PRT
; ORGANISM: Mus musculus
US-11-337-244-149

Query Match      8.4%; Score 583.5; DB 6; Length 1182;
Beet Local Similarity 24.2%; Pred. No. 3,6e-38;
Matches 282; Conservative 154; Mismatches 433; Indels 297; Gaps 44;

QY 330 SFSHTLLGQ-----FPOG---WGTWVASWPLTIVLVIPIVVALAAGLVFTTELT 376
DB 20 SAPHILHSLAPLMLRAYFOGLFSLGCRIGKKGKVLFGIVAFGLALGLRAYAIE 79
QY 377 TDPPELMSAPNSQASSEKAFHDQHG-PFRFTNQVLLTPNRSRYDSLLGPKNFSGI 435
DB 80 TDLIELMVEVSGRSVSEHLYTKKELGEBAAYTSQMLI---QTAHQEGGVLTPE---A 131
QY 436 LDLDLLELLELEOERHLQVWSPEAQNRISLQDIY---APLNPN-----TSLYDCC 486
DB 132 LD-----LHLOALTASKVOVSLYKSWDLNKKICYSKGVPIEMGIERMIEKLPFV 184
QY 487 INSLQYFQNNRNTLLLTANQTLMGTSQVWMD--HFLYCANPVL-----FKDGTALA 539
DB 185 ILTPIDCFMEGAK---LGGSAVILPERPIQWNTNDPQQLBELGFPASLEGRBLDVA 241
QY 540 LSCMADYGAPV-----PPLA---IGYKGDYS----- 565
DB 242 QVGAQVVGRCIDPDDPHCPSPAPNHSROAPVVADELSGGCHGSEHKFMHNOBELLLG 301
QY 566 -----BAVALIMTF-----SLNNYPAGDPPLAQAKUMBEAFLMEMRQFORRMAG 609
DB 302 TARDLOGQLLRBAALOSTFLMSPROLYEHFRDGYTHDIGMSEBQASVNLQWOKRFVQ 361
```

```
QY 610 MFQVTFTEKLE-----DEINRTAEDLPIPATSIYV---IFLYISIALSGSYSSWS 658
DB 362 LAQELPANAQOIHAFFSSTTLDILRAFSE-----VSTTRVVGGLTMLAAYACVMTLRWD 417
QY 659 RVMVDSKATLGLGGAVALVGLVMAAMGFYSLGRSSVLVQLVWVFLVSGADNIFLV 718
DB 418 --CAOSQAVGLAGVALLVAVSGLICALLGITFNATTQVLPFLMLGIGVDDIFLLA 475
QY 719 LEYORLPRRPGEPREVHIGRALGRVAPSMILCSSEALCFGLALTTPMAVRTALTSGL 778
DB 476 HAFKAP--PDTPLPERGECGRSTGTSVALTTSVNNVAFPMALVPIPALRAFSLQAI 533
QY 779 AVIIDFLQMSAFVALLSLDSKQESRLDYCCC-----VKQGE----- 817
DB 534 VGCNFAVMVLPFALISLDRRRHQRDLVLCFSSPCSAQVIOMLPQELGDRVAVGI 593
QY 818 -----LPP-----PG-----QGE-- 826
DB 594 AHLNATVOAFTHCEASSQHVVTILPQAHLLSPASDPLGSELVFGSGTRDLSQEBGTG 653
QY 827 -----LILGFFOKAYAPFLHMTTRGVLLPLALFGVSLYSMCHISVGLDQ 873
DB 654 PQACRPILCAHMTLAHFARVQFAPALLQTRAKALVLLFFGALLGLSLYGATLVQDGLAL 713
QY 874 ELALPKDSYLLDYFLFLNRFEVGAIVYFTVTLGYNF--SEAGMNAITCSSAGCNNSFTQ 932
DB 714 TDVVRGRTKEHAFLSAQLRYPSL-YEVALVYQGFVDYAHSGRAL-----FDLHQ 761
QY 933 KIQVATERPEOSYALIPAS-----SWVDF-----IDMLTPSSCCRLYISGPNK 976
DB 762 RFSSL-----KAVLPPTAQAPRTWLAHYRSMLOGICQAPQODNASGITHSTRNSGD 816
QY 977 D----KCPSPVNSLNCNKMCSITM-----GSVPSVEQFHKXLPWFLNDRPIKCPK 1026
DB 817 GALAVKLLIQGNQGEPLDFSQLTRKLVDEKGLIP--ELFYMGLTWVSSDPL----- 869
QY 1027 GGLA-----YST---SVNLTSDQVLAISFMAVHKPLKXSQDYTELR 1067
DB 870 -GLAASQANFYPPPEWMLHDKYDTTGEMLRIPAAQPLEFAQFPFLHGLQTAQFVEAIE 928
QY 1068 AABELAANI--TADRKVPGTDPAFEVFPYTTITNVYEQVLTILPGLFMLSIC--LYVTF 1124
DB 929 GARAACTEAGQGVAVYSGSP-----LFWEDYLG--RRCVLAVCILVCTF 976
QY 1125 AVSCLLGLDLRSGHLNLSIVMILVDVGFMAALMDISYNAVSLNVSAYGMSVEFVSH 1184
DB 977 LVYCALLLSPTWAGLI-VLVLAHMTVELFGIMFLGILKSLAPVYIIVASIGIGVEFVH 1035
QY 1185 ITRSPAISTKPTWERAKEBATISMGSAVFAGV--AMTNLPQILVGLAKAQLQIFPR 1241
DB 1036 VALGFTSHSGRNILRAA-----SALEQTFAPVTDGAVSTLGLMLAGSNPDFITRYFV 1090
QY 1242 LNLITTLGLHGLVFLVLIISYGP 1267
DB 1091 VLVTLTLLGLHGLLFLVLIISGP 1116

RESULT 13
US-60-772-265-1197
; Sequence 1197, Application US/60772265
; GENERAL INFORMATION:
; APPLICANT: BOURKHAROV, ANDREY A.
; APPLICANT: DU, ZIJIN
; APPLICANT: GUO, LIJANG
; APPLICANT: HRESKO, MICHELLE C
; APPLICANT: KOVALIC, DAVID K
; APPLICANT: ZHAOLONG, LI
; APPLICANT: LU, MAOLONG
; APPLICANT: MCCARTER, JAMES P
; APPLICANT: MILLER, NANCY M
; APPLICANT: VAUDIN, MARK
; APPLICANT: WILLIAMS, DERYCK J
```

```

? APPLICANT: WU, WEI
? TITLE OF INVENTION: IDENTIFICATION AND USE OF TARGET GENES
? TITLE OF INVENTION: FOR CONTROL OF PLANT PARASITIC NEMATODES
? FILE REFERENCE: WMDI:002USP1
? CURRENT APPLICATION NUMBER: US/60/772,265
? CURRENT FILING DATE: 2006-02-10
? NUMBER OF SEQ ID NOS: 1919
? SOFTWARE: PatentIn version 3.3
? SEQ ID NO 1197
?
? LENGTH: 891
?
? TYPE: PR1
? ORGANISM: Heterodera glycines
?
? FEATURE:
? OTHER INFORMATION: genomic DNA sequence=Seqid_396; coding sequence=Seqid_981
?
? US-60-772-265-1197

```

	Query Match	7.0%:	Score 482;	DB 8;	Length 891;		
	Best Local Similarity	22.2%:	Pred. No. 4e-30;				
	Matches	203;	Conservative	160;	Mismatches	360;	
				Indels	190;	Gaps	35;
Qy	540	LSCMADYGAVPFPFLAIGVK----	GKDYSEAEALIMTFSLNNYPAGD--PRLAQAK--L	591			
Dd	10	MNCPED-----	LIVGIRRDQAQVDIRHAEPQTVPILVAS--ACDVQRLLRSKNHR	59			
Qy	532	WEBAFLBEMRAFORRMAGMPQVTATARSLEDEINRTAADL----	PIFATS-----	639			
Dd	60	MEKEFVSFTQAAGAIIAWQRNFL--KSIDHTLNKPACQIRVNHPLASTSIQDMLEOF	117				
Qy	640	----YIVFL-YISIALGSYSWSRV-----	NWDSKATLGLOGVAUVLGAVMAMGFF	667			
Dd	118	SEFGQFVIFIGIVLMII--YAGMSQVHMQGWMFSKSCILAIIIGIVLVITLASVAGLGLS	175				
Qy	668	SYLGRSSLVLIIQVVPFLVSVGADNIIFPLEYEQRLPRRGEEREVHIGRALGRVAPSM	747				
Dd	176	TAMNHFMAATQIYIPFLTLGLGDMDMLAHNYNDVIE--AVRQEVAVALLETGMSV	232				
Qy	748	LICLSBMAICEFLGALTMPAPVRPALNSGLAVILDFLOMSAFALLSDSKROEASRL	807				
Dd	233	LTTSTINNIAFLTGCILPIPALRS CGOVALILNSVLCIILLPAFIADLRRRKXGHR	292				
Qy	808	DVCCC-----VKPELBPPPGQE-----	GLLIGFQOKYAP	838			
Dd	293	DMSFCSSRNQYAKTDSVTJTHSSIEDLQQMSAAVPTSSATKSLHKWYTLAIGFLHGYYIP	352				
Qy	839	FLLHMTTRGVVTLFLALFGVLSYMCHISVCLDELALPKDSYLDOVFLEFLNRYPEVGA	898				
Dd	353	LIRRPRIAKTVVLIYCAAMFLFCGCFGLYSRIGLELADVLPBHTPAPFLKAREXFSP-X	411				
Qy	899	PVYFV---TTLGFNFSESAENNAICSAGCNF-----	SFOKI	934			
Dd	412	PMPIYVGKPMVDY-ANQCHKLEQLARDIGRSFVIKVDGEPROMMMFMRTWLHSLSL	470				
Qy	935	OYATE--PPEOSYLAIRPASSWV-----	DPETDWLTSSCCRLYISGPNDKCPSRYN	985			
Dd	471	DRAAKIGIFPPDDCLKKMADALARNTUSDPL-LARKLIC-	SHGQ	514			
Qy	986	SLNC--LKQCMSITGMSVSRPSVEOPFKYL-PMFINDRPNIKCPKG-----	LAAYS	1033			
Dd	515	RENCETMRGAKLIEDGRINP--RGFINYLTAIFYQDNMYVYSQAAPPPTPLPMHFSAD	572				
Qy	1034	TSVNLTSDGOVASRFMAVHKPLKNSQDYTEALRAARELANITADLKRVCETDAFEVF	1093				
Dd	573	ESVVPAPD--PLYSQI.PFYMNGLTDQSIVOMIKELRAICDRYSAD-----	GLPVY	622			
Qy	1094	PYITTNVFYOXYTLIRBELFMLSLCTVTPPVSCLLDLDRSGLINLSITMI--LVD	1151				
Dd	623	PSGIPTPFMWQYIKLT---FYLSALIILIGAVLAVISITINPNMAAANAIVITWTNVE	679				
Qy	1152	TYGFMALMDISYNAVSLINLVASVNGSVEFVSHITSFAISTKPTWLERAKAETSMSGSA	1211				
Dd	680	LAFMGVGFVKOMPISAVILTITAVGIVEFTAVVLAPLTS-----LGRDERMVACIEH	734				
Qy	1212	VPRGV--AMTNLPGLIVLGLARQOLIQIFFPRLNLITLTLGLHGLVPLPYILSVYG-	1267				

[illegible]

```

RESULT 14
US-60-772-265-241
/ Sequence 241, Application US/60772265
/ GENERAL INFORMATION:
/ APPLICANT: BOURKHAROV, ANDREY A.
/ APPLICANT: DU, ZIJIN
/ APPLICANT: GUO, LIANG
/ APPLICANT: HRESKO, MICHELLE C
/ APPLICANT: KOVALIC, DAVID K
/ APPLICANT: ZHAOLONG, LI
/ APPLICANT: LU, MAOLONG
/ APPLICANT: MCCARTER, JAMES P
/ APPLICANT: MILLER, NANCY M
/ APPLICANT: VAUDIN, MARK
/ APPLICANT: WILLIAMS, DERYCK J
/ APPLICANT: WU, WEI
/ TITLE OF INVENTION: IDENTIFICATION AND USE OF TARGET GENES
/ TITLE OF INVENTION: FOR CONTROL OF PLANT PARASITIC NEMATODES
/ FILE REFERENCE: NMID:002USP1
/ CURRENT APPLICATION NUMBER: US/60/772, 265
/ CURRENT FILING DATE: 2006-02-10
/ NUMBER OF SEQ ID NOS: 1919
/ SOFTWARE: PatentIn version 3.3
/ SEQ ID NO 241
/ LENGTH: 1358
/ TYPE: PRT
/ ORGANISM: Caenorhabditis elegans
/ FEATURE:
/ OTHER INFORMATION: C elegans public gene name=Y110A2AL.8
/ FEATURE:
/ OTHER INFORMATION: genomic DNA homolog sequence=SeqID_396
US-60-772-265-241

```

Query Match	6.6%	Score 455;	DB 8;	Length 1358;
Best Local Similarity	22.8%	Pred. No. 1.2e-27;		
Matches 215;	Conservative 146;	Mismatches 373;	Indels 209;	Gaps 37;
Qy	473	APLNPNTSLYQCCI--NSLIQYQNNRTLLLTANOTLNGQTSQVMDKHFLYCANAP	529	
Db	449	AKKNPEKAKDILCLEYSSSLKKMMQENPERL-----GEFLTEKEMP	490	
Qy	530	LTFKQGTALASQMDYGAPEPFP--LAIGYK--GKDYSSEALIMTESLANNYPAGD	583	
Db	491	DYPNYGDVMTGGCKG-FGKKIMWEDEDLIGIGIQDNGCLYS-AALLOSVELVSGAYDV	548	
Qy	584	PRLAQKLMBEAFLFEMRAFORRMG---MEQVFTARSLEDEINRTTAEDLPFATS	639	
Db	549	ARIKNDKTDHPGLDR-HHFQPMAGEIISTQYRNF-KRLYSHLNNRRRQFPHLASTS	606	
Qy	640	-----YIVFLYLSIALGSYSWSRWYDSKATLGLGVAVYLVAV	680	
Db	607	IADMLEEFSQFNIIIVIGIILMVIYAALFQGRFQGM-WLAQSNVALAICGVIIIVTSS	665	
Qy	681	MAAMGFPSYIGTRSSLVIIQVYPPFLVLSGADNIFIFVLEYQRLPRPGEPREVHIGRL	740	
Db	666	ICGLGEATMIGINFNATTOVVPFLSLGIGIDMFLLNAYBIINICKN--BIGVLL	722	
Qy	741	GRAPASMLCSLEAICFFLGALTPEMPARTEPALTSGIAVIIDPFLQNSAFVALLSLDK	800	
Db	723	KETGMSVMLISINNIIAIFISGYLPLPALRFSQCAIILIAFLNFIPLMFIPLFAMIGIDLR	782	


```
QY 801 ROEASRLDYCCVK-----POELP-----PPGCEGL 828
      : : : : :
Db 783 RORAKRIDLAICSRGPNQWATSQVPSVSNSSRAELAGYEKQADEYKRHEFWYVG-- 840
QY 829 LGFPGKAVAPFLHMTTRGVLLFLALFGVSLYSCHISVGLDQELAPKXSYLLDYFL 888
      : : : : :
Db 841 -GFANKIYIPALKNNVAKCVLIGTTTAVVFGIYKMYSTLGLADLVIPKTPPAFIR 899
QY 889 FLNRYFEVGAPVYFV-----TTLGYNFSSEAGNNAICSSAGNCSFPTOKIQYA 937
      : : : : :
Db 900 AEEQYFSF-YPMFAVLRGDKLDIPNQOQILEYBAQLG-----SSKFMIX 943
QY 938 TEFPESOSTAIAPASSVVDPIFML-----TPSS- 965
      : : : : :
Db 944 AEGKLQPY-----WMSMLRYWLSGLDMALEKDLAAGKFDLTNGNPIKNGEKPESPM 996
QY 966 -CCRLYIS-GPNKDKCPSTVNSLNLCKMCSITMGSVRPSVEQPHLYP-WFLNDRPMI 1022
      : : : : :
Db 997 IAAKLYCSFCTYN--CDORLKKMKVENEV-----INP--EGFYNYLTGMFNVD--NM 1044
QY 1023 KCPKGLAAYSTSVNLTSDOVLASRFMAVHKPLKNSQ-----DYTEALRAARELA 1073
      : : : : :
Db 1045 MYVQASAFYTPRPGEVNEKL--AKVPAABEFLYSQMPFYQNDLIDTPALVKMIEIR 1102
QY 1074 ANITDLRKVPCTDPAFEVFPYTTINVFYEQYITLPEGLFMLSCLVETTFVASCLLGL 1133
      : : : : :
Db 1103 ATCE-----EYSEKLSNHPGSIAPTFWEQYLT-LRMNLFQ-ALAVFCVISTL 1153
QY 1134 DLRSGLNLSIVML--VDTVGFMAIMDISYNAVSLINLVSAVGSVEFVSHITSPAI 1191
      : : : : :
Db 1154 MENPNAATILKCIIVITTELGEFGMLKGIKNMPISAVTLICAVGIVGFTHAVEIAF-L 1212
QY 1192 STKPTMLEBAKATISMGSAVFAGVAMTNLPGLIVLGLAKAQLIQIFFEFLNLLITLGL 1251
      : : : : :
Db 1213 TALGTHDQLBESCLQMFVPVHGAISTFL-GVMVLFSEPDFVVFYFTMTLVALGV 1271
QY 1252 LHGLVFLPYILSYGPD-----VNPALALEQKAEBA 1283
      : : : : :
Db 1272 FNLGLCVLPVILTLVGPKEPLTPTDSSVLPPPPFLQYAEKS 1314

RESULT 15
US-11-360-355-133805
; Sequence 133805, Application US/11360355
; GENERAL INFORMATION:
; APPLICANT: Boukharov, Andrey
; APPLICANT: Du, Zifang
; APPLICANT: Guo, Liang
; APPLICANT: Kovalic, David
; APPLICANT: Lu, Maolong
; APPLICANT: McCarter, James
; APPLICANT: Miller, Nancy
; APPLICANT: Williams, Deryck
; APPLICANT: Vaudin, Mark
; APPLICANT: Wu, Wei
; TITLE OF INVENTION: METHODS FOR GENETIC CONTROL OF HETERODERA INFESTATIONS
; FILE REFERENCE: IN PLANTS AND COMPOSITIONS THEREOF
; CURRENT APPLICATION NUMBER: US/11/360,355
; CURRENT FILING DATE: 2006-02-24
; NUMBER OF SEQ ID NOS: 171306
; SEQ ID NO 133805
; LENGTH: 831
; TYPE: PRP
; ORGANISM: Heterodera glycines
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (794)..(794)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
; FEATURE:
; OTHER INFORMATION: Coding regions on vcdna: vcdna=SeqID 60328; Strand=-; Position=1
; OTHER INFORMATION: -173,414-469,516-626,687-921,1040-1216,1282-1406,1470-1579,1627
; OTHER INFORMATION: -1785,1833-1933,1984-2145,2192-2241,2295-2457,2503-2695,2750
```

```
; OTHER INFORMATION: -2985,3164-3208
; FEATURE:
; OTHER INFORMATION: Homolog annotation: Hlt_ID=NP_496761.1; Match level="QueryCover
; OTHER INFORMATION: =99%, HltCov=99%, E-value=0.0, Identity=48%", Hlt descrip
; OTHER INFORMATION: =Patched Related (90.8 kd) [pct=18] [Caenorhabditis elegans]
; OTHER INFORMATION: p1r|126683 hypothetical protein Y38F1A.3 - Cae
US-11-360-355-133805
```

```
Query Match 6.3%; Score 437; DB 7; Length 831;
Best Local Similarity 22.5%; Pred. No. 1,6e-26;
Matches 189; Conservative 139; Mismatches 303; Indels 210; Gaps 30;
```

```
QY 590 KMEBAFLEEMAFQRRAGMGQVYFTIERSLEDSINTTADLIIFATSYVIFLYISL 649
      : : : : :
Db 21 KMEBSVYN--FAMNSRGDPLKHLVTSBGVSEVRRTGLEVLPLMPISLVIIIFYI 78
QY 650 A-----LGSY-----SSWSRVWVDSKATGLGCVAVVLGAVMAAGFSY 689
      : : : : :
Db 79 TSLKRDQIRSKPMELVGVFCPTVITSLKRDQIRSKPMELVGVFCPLISVGASFGTLFW 138
QY 690 LGIR-----SGLVILQVVPFLVLSVGADNIFIVLEYQRLPRRGEPREVHIGRALGV 743
      : : : : :
Db 139 LGFBEFLPGFBEFLPILVVVFLIALIGVDDVIFLHCWHTD--PSKOLREKRVADLLGSA 196
QY 744 APSMLCSISEAICFFLGALTPMAVRTPALTSGLAVIDFLQMSAFALLSLSKQOE 803
      : : : : :
Db 197 GPSVITTSITLWMLSTFIATPTPAIRTFCLFISVAIVAYLYOLFPTAVAVVIGAORBA 256
QY 804 ASRLDYCCVK--POELPPQG-----GEGLLGFGQKAYAPFLHWM 843
      : : : : :
Db 257 DERNAYPLCIKVPKPSRPVATLKKETNLMRLGSTNMLRSLSDV----DRYVDFWWSW 311
QY 844 ITR---GVVLLLEFLAL--FGVSLYSCHISVGLDQELAPKDSYLLDYFLNLR--YFEV 896
      : : : : :
Db 312 TARIILASLIITYMAFSAVGAQIKVQIKVGLTSEKLFELDDSPLELVRLONNVYFKEG 371
QY 897 GAVPYFTTTLGYNFSSBAGMNAICSSAG--CNPSTOKIQVATEPPEOSTYLAIPASS--- 952
      : : : : :
Db 372 GQMAVFFVNSPG--DLRKMAVFFVNSPGDLRKPSAIPIKIMLLEBFESQNSGVSSTOM 428
QY 953 WYDDFIDMLTPSSCCRLYISGPNKDKCPSTVNSLNLCKMCSITMGSVR-----P 1003
      : : : : :
Db 429 WINTYLPFIG-----LONRGSINRGSIDFRYKTYLDFP 461
QY 1004 SVEQPHLYLPMF-----LNDRENIKCPKGLAAYSTSVNLTSDOVLASRFMAVHKP 1055
      : : : : :
Db 462 SIPEYHRMHSFVSLGPKEDCLNERP-----SCINKFPDCLNERBSCINKF--FSTG 511
QY 1056 LKNS--QDYTEALRAARELA-----NITADLRKVPCTDPAFEVFPYTTINVFYEQYIT 1107
      : : : : :
Db 512 FQNAVAMSDELVLQRMWQLADYSQNLIT-----YVEDFSMAVDQADQLLS 560
QY 1108 I-----LPEGLFMLSCLVETTFVASCLLGL-----DLRSGLNLSIVMLVDTVGFPMAL 1158
      : : : : :
Db 561 IPSVTIIQVAPFALLCMTFVAFALLCMTFVLMVFTPSISITLPGTACVLSINLGVGILFY 620
QY 1159 WDISYNAVSLINLVSAVGSVEFVSHI-----TSFAISTKPTWL----- 1198
      : : : : :
Db 621 WSIDIDPISMTTTLMAIGSVDFVAHISPHYKGMTTTLMAIGSVDFVAHISPHYKGE 680
QY 1199 -----ERAKEATISMGSAVFAGVAMTNLPGLIVLGLAKAQLIQIF----- 1238
      : : : : :
Db 681 IEDSERERLHALSSIAMPLQ--AALSTVLSLMIIVLIHAYWQVAVKVVLLVVVLGLVHG 739
QY 1239 ---FFRINLLITLLGLVFLPYLSVY-----GPDVN-----ALALEOKRA 1280
      : : : : :
Db 740 LVFVKVVLVVLVGLVHGLVLPVVAIIPQKSGATKAPKVPPLNRRRTIYTXOKRRR 799
QY 1281 E 1281
      : : : : :
Db 800 E 800
```

Search completed: March 22, 2006, 23:07:41
Job time : 63 secs

THIS PAGE BLANK (uspto)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2006, 23:07:52 ; Search time 25 Seconds
(without alignments)
1525.028 Million cell updates/sec

Title: US-10-736-769-4

Sequence: 1 MABAGRGWMLWMLRLRLAQ.....GSIKGAGAIINFLPNNRGPF 1332

Scoring table:

BLOSUM62

Searched: 169630 seqs, 28622889 residues

Total number of hits satisfying chosen parameters: 169630

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: Published Applications_AA_New:*
2: /cgn2_6/ptodata/1/pubppaa/US08_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubppaa/US07_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubppaa/US09_NEW_PUB.pep:*
5: /cgn2_6/ptodata/1/pubppaa/US10_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubppaa/US11_NEW_PUB.pep:*
7: /cgn2_6/ptodata/1/pubppaa/US12_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubppaa/US60_NEW_PUB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6896	99.8	1332	US-11-242-459-9	Sequence 9, Appl1
2	2402.5	34.8	1278	US-10-995-561-952	Sequence 952, App
3	942	13.6	194	US-11-242-459-8	Sequence 8, Appl1
4	195.5	2.8	1087	US-11-242-459-17	Sequence 17, Appl1
5	195.5	2.8	1392	US-11-242-459-40	Sequence 40, Appl1
6	191	2.8	127	US-11-096-568A-6778	Sequence 6778, Ap
7	173	2.5	85	US-11-096-568A-6779	Sequence 6779, Ap
8	143	2.1	445	US-11-242-459-35	Sequence 35, Appl1
9	143	2.1	456	US-11-242-459-34	Sequence 34, Appl1
10	141.5	2.0	1048	US-10-392-234A-18	Sequence 18, Appl1
11	140.5	2.0	1046	US-10-392-234A-16	Sequence 16, Appl1
12	134	1.9	1067	US-10-467-657-2648	Sequence 2648, Ap
13	127	1.8	1048	US-10-392-234A-14	Sequence 14, Appl1
14	126.5	1.8	200	US-11-242-459-16	Sequence 16, Appl1
15	126	1.8	801	US-10-793-626-2020	Sequence 2020, Ap
16	121.5	1.8	1043	US-10-392-234A-34	Sequence 34, Appl1
17	120.5	1.7	2426	US-11-203-806A-11	Sequence 11, Appl1
18	119	1.7	1049	US-10-392-234A-20	Sequence 20, Appl1
19	117.5	1.7	1049	US-10-392-234A-12	Sequence 12, Appl1
20	115.5	1.7	914	US-11-072-512-2923	Sequence 2923, Ap
21	113	1.6	1034	US-10-392-234A-30	Sequence 30, Appl1
22	112.5	1.6	697	US-11-082-389-362	Sequence 362, App
23	109.5	1.6	758	US-11-096-568A-30412	Sequence 30412, A
24	109.5	1.6	772	US-11-096-568A-30411	Sequence 30411, A
25	109.5	1.6	827	US-11-096-568A-30410	Sequence 30410, A

26	109.5	1.6	1049	7	US-11-098-686-11239	Sequence 11239, A
27	109	1.6	1032	6	US-10-392-234A-67	Sequence 67, Appl1
28	108.5	1.6	648	7	US-11-087-099-4135	Sequence 4135, Ap
29	106	1.5	443	7	US-11-087-099-11457	Sequence 11457, A
30	105.5	1.5	783	7	US-11-082-389-354	Sequence 354, App
31	105	1.5	324	7	US-11-087-099-11792	Sequence 11792, A
32	105	1.5	1080	7	US-11-096-568A-27723	Sequence 27723, A
33	105	1.5	1097	7	US-11-096-568A-27722	Sequence 27722, A
34	105	1.5	1128	7	US-11-096-568A-27721	Sequence 27721, A
35	104.5	1.5	315	7	US-11-087-099-9820	Sequence 9820, Ap
36	104.5	1.5	471	7	US-11-087-099-4362	Sequence 4362, Ap
37	104	1.5	757	7	US-11-096-568A-27509	Sequence 27509, A
38	104	1.5	771	7	US-11-096-568A-27508	Sequence 27508, A
39	104	1.5	826	7	US-11-096-568A-27507	Sequence 27507, A
40	103.5	1.5	251	7	US-11-087-099-1098	Sequence 1098, Ap
41	103.5	1.5	331	7	US-11-087-099-8831	Sequence 8831, Appl
42	103.5	1.5	2312	7	US-11-126-313-34	Sequence 34, Appl
43	103	1.5	463	7	US-11-087-099-3980	Sequence 3980, Ap
44	102.5	1.5	475	7	US-11-087-099-2723	Sequence 2723, Ap
45	102	1.5	324	7	US-11-087-099-5717	Sequence 5717, Ap

ALIGNMENTS

RESULT 1									
US-11-242-459-9									
Sequence 9, Application US/11242459									
Publication No. US20060035835A1									
GENERAL INFORMATION:									
APPLICANT: TANIYAMA, Yoshio									
APPLICANT: KITTA, Shunbun									
TITLE OF INVENTION: Novel Protein, Process for Producing The Same And Use Thereof									
FILE REFERENCE: 270US06									
CURRENT APPLICATION NUMBER: US/11/242,459									
CURRENT FILING DATE: 2005-10-03									
PRIOR APPLICATION NUMBER: US/10/239,316									
PRIOR FILING DATE: 2002-09-19									
PRIOR APPLICATION NUMBER: PCT/JP01/02279									
PRIOR FILING DATE: 2001-03-22									
PRIOR APPLICATION NUMBER: JP2000-088595									
PRIOR FILING DATE: 2000-03-24									
NUMBER OF SEQ ID NOS: 59									
SEQ ID NO 9									
LENGTH: 1332									
TYPE: PRT									
ORGANISM: Human									
US-11-242-459-9									
Query Match									
Best Local Similarity 99.8%; Score 6896; DB 7; Length 1332;									
Matches 1330; Conservative 0; Mismatches 2; Indels 0; Gaps 0;									
QY	1	MABAGRGWMLWMLRLRLAQSEPYTTHQGYCAFYDEGCKNPELSGLMTLSNVCLSN	60						
DB	1	MABAGRGWMLWMLRLRLAQSEPYTTHQGYCAFYDEGCKNPELSGLMTLSNVCLSN	60						
QY	61	TPARKITGDHLLILOKICPRLYGPNTQACCAKQVLSIASISTKALITRCPACSDNF	120						
DB	61	TPARKITGDHLLILOKICPRLYGPNTQACCAKQVLSIASISTKALITRCPACSDNF	120						
QY	121	VNIHCNHTSPNOSTLFTNTRVQGLAGOLPAVVAEAFQHSFAROSYDCSRVVPAA	180						
DB	121	VNIHCNHTSPNOSTLFTNTRVQGLAGOLPAVVAEAFQHSFAROSYDCSRVVPAA	180						
QY	181	ATLAVGTMCGVYSALCNARWLNFGODTGNGLAPDITFHLPEQAVSGIOPINEGV	240						
DB	181	ATLAVGTMCGVYSALCNARWLNFGODTGNGLAPDITFHLPEQAVSGIOPINEGV	240						
QY	241	ARCNEGQDDVATVCSQDCDCAACPAIARPOLDSTYTLGQMPSTVLIIILCSVPAVVT	300						
DB	241	ARCNEGQDDVATVCSQDCDCAACPAIARPOLDSTYTLGQMPSTVLIIILCSVPAVVT	300						

QY 301 LLVGFVAPADPKSKXNDPKKSTLSLSDKLSFSTHTLLGQFPQGMGTWASWPLTILVLSV 360
DB 301 LLVGFVAPADPKSKXNDPKKSTLSLSDKLSFSTHTLLGQFPQGMGTWASWPLTILVLSV 360
QY 361 IPVVALLAAGLVTELTTPDVELMSAPNSQARSSEKADHDFGFPPTNOVITLAPRRSY 420
DB 361 IPVVALLAAGLVTELTTPDVELMSAPNSQARSSEKADHDFGFPPTNOVITLAPRRSY 420
QY 421 RYDSSLGPKNFGSLIDLDLLELLELOERLRLQVMSPEAQRNLSLODICVAPLNDPT 480
DB 421 RYDSSLGPKNFGSLIDLDLLELLELOERLRLQVMSPEAQRNLSLODICVAPLNDPT 480
QY 481 SLYDCCINSLLQYFQNNRTLLLTITANQTMGQTSQVDMKDHFLYCANAPLTFKQGTALAL 540
DB 481 SLYDCCINSLLQYFQNNRTLLLTITANQTMGQTSQVDMKDHFLYCANAPLTFKQGTALAL 540
QY 541 SCMDVGAVFPPRLAIGYKGDYSEAEALIMTFSLNNYPAQDPRLAQAKLMEAEALDEM 600
DB 541 SCMDVGAVFPPRLAIGYKGDYSEAEALIMTFSLNNYPAQDPRLAQAKLMEAEALDEM 600
QY 601 RAFOREMAGFOVTFPAERSLEDEINRTAEDLPFATSYIVFLYISLALGYSMSRY 660
DB 601 RAFOREMAGFOVTFPAERSLEDEINRTAEDLPFATSYIVFLYISLALGYSMSRY 660
QY 661 MYDSKATLIGVAVVLAGVMAAMGFSSYLGRSSLVIIQVVPFLVSVGADNIFIVLE 720
DB 661 MYDSKATLIGVAVVLAGVMAAMGFSSYLGRSSLVIIQVVPFLVSVGADNIFIVLE 720
QY 721 YQRLPRRBPBPBREVHGRALGRVABSMILCSLEALCFPLGALTMPNARTPALTSGLAV 780
DB 721 YQRLPRRBPBPBREVHGRALGRVABSMILCSLEALCFPLGALTMPNARTPALTSGLAV 780
QY 781 IIDFLQMSAFVALLSDSKROBASRLDYCCCVKPOELPPGQSEGLLGFQKAYAPPL 840
DB 781 IIDFLQMSAFVALLSDSKROBASRLDYCCCVKPOELPPGQSEGLLGFQKAYAPPL 840
QY 841 LHMTRGVALLFLALFGVSLYSWCHI SVGLDQELAPKDSYLLDYFLFLNRYFEVGA 900
DB 841 LHMTRGVALLFLALFGVSLYSWCHI SVGLDQELAPKDSYLLDYFLFLNRYFEVGA 900
QY 901 YPFTTLGVNFSBAGNNALCSSAGCNNSFTQKIQVATEPBPESYLAIPASSVVDPIFM 960
DB 901 YPFTTLGVNFSBAGNNALCSSAGCNNSFTQKIQVATEPBPESYLAIPASSVVDPIFM 960
QY 961 LTPSSCCRLYISGPNKDFCPTSVNSLNCIKNCSITMGSVSPSVQFHKYLPWFLNDRP 1020
DB 961 LTPSSCCRLYISGPNKDFCPTSVNSLNCIKNCSITMGSVSPSVQFHKYLPWFLNDRP 1020
QY 1021 NIKCPKGLAAVYSTVNLTSQGVLASRFMAVHKPKNSQDYTEALRAARELANITADL 1080
DB 1021 NIKCPKGLAAVYSTVNLTSQGVLASRFMAVHKPKNSQDYTEALRAARELANITADL 1080
QY 1081 RKPFGDPAPEVFPYITTVFEOYLITLPEGLFMLSCLVPFPAVSCLLGLDLSGLL 1140
DB 1081 RKPFGDPAPEVFPYITTVFEOYLITLPEGLFMLSCLVPFPAVSCLLGLDLSGLL 1140
QY 1141 NLLSIWMLVDYTGFAALMDISYNAVSLNLVSAVGMSVEFVSHIRSFALSTKPTWBLR 1200
DB 1141 NLLSIWMLVDYTGFAALMDISYNAVSLNLVSAVGMSVEFVSHIRSFALSTKPTWBLR 1200
QY 1201 AKEATISMGSAVFAGVAMTNLPGILVGLAKAQLIIFPERLMLITLLGLLHGLVFLPV 1260
DB 1201 AKEATISMGSAVFAGVAMTNLPGILVGLAKAQLIIFPERLMLITLLGLLHGLVFLPV 1260
QY 1261 ILSYVPDVPNPAALAQKABEAVAAVMVASCNHPSSRVSTADNIVNHSFBSIGAGA 1320
DB 1261 ILSYVPDVPNPAALAQKABEAVAAVMVASCNHPSSRVSTADNIVNHSFBSIGAGA 1320
QY 1321 ISNFLPNNGRQF 1332
DB 1321 ISNFLPNNGRQF 1332

RESULT 2
US-10-995-561-952
; Sequence 952, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CLO01559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FaastSeq for windows version 4.0
; SEQ ID NO 952
; LENGTH: 1278
; TYPE: PRF
; ORGANISM: Homo sapiens
US-10-995-561-952

Query Match 34.8%; Score 2402.5; DB 6; Length 1278;
Best Local Similarity 39.8%; Pred. No. 2,6e-181;
Matches 522; Conservative 237; Mismatches 450; Indels 103; Gaps 25;

QY 7 RGMILLALLRLAQEPYTTIHQPGYCAFYDECG-----KNPELSGLMTLSNVSCLS 59
DB 4 RLALGLLTLTLCPAQVFSQ-----SCWYGCGLAYGDKRNCVSG----- 46

QY 60 NTPAKITGDDHLLLOKICPRLYTGPNTQACCSAQVLSLEASLITKALLTRCPACSDN 119
DB 47 --PPPLPDGDYDLVQELCPGFFFG--NVSLLCCDVQLOTLKDNLOLPQFJSRCSFCRYN 103

QY 120 FVNLCHNTCSPNQSLFNTVTR----VAQLGAGQLPAVAAEAFQHSFAESQYDSCSRV 175
DB 104 LNLCELTCSRQSGFLNVTATEDYVDVPTNQTNTVKEIQYVYGQGFANMVAACDV 163

QY 176 RVPAAATLAVTGCVGSALCNAQRWLNFGQDGTNGALPLDIT-----PHLLPQQA 228
DB 164 EAPSSNDKALGLLCKGKDA--CNATNMIEWMKNQNGAAPFTITVFSDFVH----- 215

QY 225 VSSGIQPLNEGVARBNESQGDVATCSQDCAACGPAIARPO-----ALDSTFYLG 279
DB 216 --GMEPMNATKXGCDSEVDEVTAPSCQDCSIVGPKRPDPAPPAPMTILGLDMYIM 272

QY 280 QMPGSLVLIILCSYFAVVTILL-----VGFVAPADPKSKMVDPKKSTLS 326
DB 273 WITWMAFLVFGAFPAFWCVRKRYFVSEYTPIDSNIAFSV--NASDKGE-----ASCC 324

QY 327 DKLSESTHTLLGQFPQGMGTWASWPLTILVLSVIVVALLAGLVTELTTPDVELMSAP 386
DB 325 DVSAAFAEGCLRLFTBRGSCFCVRNPGCVIIFPSLVIITACSGLGFVVTTPVPLWSAP 384

QY 387 NQOASSEKAFHQHGHGPPFRITNOVILTPANMSRYDSDLLGPK--NFGSIIDLDLLELL 445
DB 385 SSQARLEKEYFQHGHPFRTEQLITRAPLTDKHIQVPPSGADVPPPPDIQILHVL 444

QY 446 ELQERLRLQVMSPEAQRNLSLODICVAPLNDPTSLDCCINSLLQYFQNNRTLLLELTA 505
DB 445 DLQIAIEN--ITASDNETVITQDICLAPLSYNT--NCTILSYLANTFQNSHSLDKK 499

QY 506 NOTLMGQTSQVDMKDHFLYCANAPLTFKQGTALALSCHADYGAPVFPPLAIGYKGDYS 565
DB 500 GDDPF---VYADYHHPFLYCVAPASLNDTSLHDPCLGTGCGPVFPMLVIGGYDDQYN 556

QY 566 EAEALIMTFSLNNYPAQDPRLAQAKLMEAEALFEKRAQORMAAGFOVTFPAERSLEDEI 625
DB 557 NATAVITFPVNNYNDTEKQRAQAWKEKFNFPVKNK--PMLTISFPAERSIEBEL 613

QY 626 NRTTAEDLPFATSYIVFLYISLALGYSMSRWVMSKATLIGVAVVLAGVMAAMG 685
DB 614 NRESDSVFTVVISITAFIYISLALGHIKSCRLLVDSKSLGAGIILVLSVACSLG 673

QY 686 PRSYLGRSSVLVILQVPPVLVSVGADNIFIVLEVOQLPRRPREVHIGALGRVAP 745
 DB 674 VESYGLPLPLVILVIEVILPFLVAVGVNDIFILVOAYORDERLQGETLDDQLGRVLEBPV 733
 QY 746 SMILGSLSEALCFGLATPMPAVRTFALTSGLAVIDLLOMSAPALISLSKOEAS 805
 DB 734 SMFLSPSEVAVFALGALSWPAVHTFSLPAGLAVFDLQITCFVSLGIDIKOEKN 793
 QY 806 RLDDVCCVPEQLPPEPGQ-GEGLLGFQKAVAPFLHMTTRGVVLLLFALFGVSLYSW 864
 DB 794 RLDDIFCCVGAELGTSVQASESCLFRPFKNSYSPFLIKMMRPVIAIFGVLSFSLAVL 853
 QY 865 CHISVGLDELALPKDSYLLDYFLFNRYFEGVAPYFVTLTGYNFSSSEAGMNAICSSAG 924
 DB 854 NKVDIGLDQSLSMDDSYWVDYFKSISQYLHAGRPYFVLEBHDYTSKGMVWCGMG 913
 QY 925 CNNSFTOKIQVATEEPPEGSYLAIRASSWDDFDMLTP-SGCCLYISGPNKDCPST 993
 DB 914 CNNSDLVQOIFNAQOLDNRYTRIGFADSSWIDYFDWVKPQSSCCRY--DNITDQCNAS 970
 QY 984 VNSLNCXKCMST-MGSVSPVEQFHXYLPFLNDRPNKCPKGGALASTSVN--LTS 1040
 DB 971 VVDPACVR-CRPTPEKQKPOGGDFMRFLPMLSDNPNKCGKGAHAAVSAVNLIGH 1029
 QY 1041 DQOVLASRPMAYHKPKNSQDYTEALRAARELANITADLRKVPGTDPAFVFPYTTNV 1100
 DB 1030 GRFVGATVEMTYHTVLTQTSADPFDALKKALIASNT-ETMGINGS--AVRVPYSVYV 1086
 QY 1101 FPEQVITLPEGLFMSLCLVPPFVNSCLLGLDLSGLNLSTYVILVDTYGFALMD 1160
 DB 1087 FPEQVITLIDDTIFNLGVSLGALFLVYVLLGCELSAVIMCATIAMVLMNMGMVLMWG 1146
 QY 1161 ISYNAVSLNLVSGVSGVEFVSHITRSPALSTFPTLEBAKATISMGSAVAGVAMTN 1220
 DB 1147 ISLNAVSLVNLVWSCGISVEFCSHITRAFVSKGSAVEAEHLAMGSSVSGITLTK 1206
 QY 1221 LPGILVGLAKAQLIQIFFRMLNLITLLGLHGLVFLPVILSVYGPVNP 1272
 DB 1207 FGIIVVLAFAKSGIQIFVYFRMYLAVLIGATHGLFLPLVLSYIGPSVVK 1258

RESULT 3
 US-11-242-459-8
 ; Sequence 8, Application US/11242459
 ; Publication No. US20060035835A1
 ; GENERAL INFORMATION:
 ; APPLICANT: TANUYAMA, Yoshio
 ; APPLICANT: KITA, Shunbun
 ; APPLICANT: SATOMI, Tomoko Komiya
 ; TITLE OF INVENTION: Novel Protein, Process for Producing The Same And Use Thereof
 ; FILE REFERENCE: 2703USOP
 ; CURRENT APPLICATION NUMBER: US/11/242,459
 ; PRIOR FILING DATE: 2005-10-03
 ; PRIOR APPLICATION NUMBER: US/10/239,316
 ; PRIOR FILING DATE: 2002-09-19
 ; PRIOR APPLICATION NUMBER: PCT/JP01/022279
 ; PRIOR FILING DATE: 2001-03-22
 ; PRIOR APPLICATION NUMBER: JP2000-088595
 ; PRIOR FILING DATE: 2000-03-24
 ; NUMBER OF SEQ ID NOS: 59
 ; SEQ ID NO 8
 ; LENGTH: 194
 ; TYPE: PRT
 ; ORGANISM: Human
 ; US-11-242-459-8

Query Match 13.6%; Score 942; DB 7; Length 194;
 Best Local Similarity 99.5%; Pred. No. 4.8e-67;
 Matches 193; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 615 FTMERSLEDEINNTTADLPFIATSYIVFLYISLALGSSWSRWVWDSKATLGLGVA 674
 DB 1 FMAERSLDEINNTTADLPFIATSYIVFLYISLALGSSWSRWVWDSKATLGLGVA 60

QY 675 VVLGAVMAAMGFPSYLGIRSSVLVILQVPPVLVSVGADNIFIVLEVOQLPRRPREV 734
 DB 61 VVLGAVMAAMGFPSYLGIRSSVLVILQVPPVLVSVGADNIFIVLEVOQLPRRPREV 120
 QY 735 HIGRALGRVAPSMILGSLSEALCFGLATPMPAVRTFALTSGLAVIDLLOMSAPV 794
 DB 121 HIGRALGRVAPSMILGSLSEALCFGLATPMPAVRTFALTSGLAVIDLLOMSAPV 180
 QY 795 LSLDSKROEASRLD 808
 DB 181 LSLDSKROEASRLD 194

RESULT 4
 US-11-242-459-17
 ; Sequence 17, Application US/11242459
 ; Publication No. US20060035835A1
 ; GENERAL INFORMATION:
 ; APPLICANT: TANUYAMA, Yoshio
 ; APPLICANT: KITA, Shunbun
 ; APPLICANT: SATOMI, Tomoko Komiya
 ; TITLE OF INVENTION: Novel Protein, Process for Producing The Same And Use Thereof
 ; FILE REFERENCE: 2703USOP
 ; CURRENT APPLICATION NUMBER: US/11/242,459
 ; PRIOR FILING DATE: 2005-10-03
 ; PRIOR APPLICATION NUMBER: US/10/239,316
 ; PRIOR FILING DATE: 2002-09-19
 ; PRIOR APPLICATION NUMBER: PCT/JP01/022279
 ; PRIOR FILING DATE: 2001-03-22
 ; PRIOR APPLICATION NUMBER: JP2000-088595
 ; NUMBER OF SEQ ID NOS: 59
 ; SEQ ID NO 17
 ; LENGTH: 1087
 ; TYPE: PRT
 ; ORGANISM: Human
 ; US-11-242-459-17

Query Match 2.8%; Score 195.5; DB 7; Length 1087;
 Best Local Similarity 19.5%; Pred. No. 5.5e-07;
 Matches 200; Conservative 129; Mismatches 320; Indels 379; Gaps 47;
 QY 576 LNNYPAGDPRLAQAKMEBAFLEMBRAFQRMAGMFOVTFARSLD----- 623
 DB 100 LPNYVSVDLR-----WEB-----QRAKQSFVTVYVAMLANQSTKVQVYGSTD 144
 QY 624 ----EINRTTAD--LPFATSYIVFLYISLALGSSWSRWVWDSKATLGLG--VAVV 676
 DB 145 LPDYEVARTFNNMDLAFISSSCIALVYILTSCVFLSFFGI-----ASIGLSCLVALP 199
 QY 677 LGAVMAAMGFPSYLGIRSSVLVILQVPPVLVSVGADNIFIVLEVOQLPRRPREV 735
 DB 200 LHVHVFGL--QVLGLING-----VAFVYVIGVDVDFVINTYQATHLEDPQLBMR 251
 QY 736 IGRALGRVAPSMILGSLSEALCFGLATPMPAVRTFALTSGLAVIDLLOMSAPV 792
 DB 252 TVQGTAK-----ATFSTLTTRAAVAAVVFQIPAVHFGFLMSLIVGCCMLAVLVTMPAL 308
 QY 793 ALISLDSKROEASRLDVC--CCYKQDELPPPGQ----- 823
 DB 309 GLWSLYLADPLESSCQTSCHQNCRSKTSLHPGDPVAPAEVGSPPAQGPVPLDDDIPL 368
 QY 824 -----GEGLLGFQKAVAP-----FLHMI-----TRGVVL 850
 DB 369 EYEEPEVSLBELGVSLVSVSPBGLQPASTGSRGHLIVQLQELHHVLMASVAKSKWIV 428
 QY 851 LFLFALFGVSL-----YSM-----CHISVGLDOE- 874
 DB 429 GLFVSLILISLVAASRLRPASRAPLFRPDNTIQVLLDKYNSASAGISCTTSGLFQEK 488
 QY 875 -----LALPK-----DSYLLDY--FLFNRYFEGVAPYFVTLTGYNP 910

Db 489 PHSLONNIRTSLEKRRGSGVPMASRPATLQDPCTVYISKYKSGHRA--VYRLSLNA 546
Qy 911 SSEAGMNAICSSAG--CNNF-----SFTOKIQVATEPPEOSYLAIPASWVDDFIML 961
Db 547 SLPAWQAVSPGDGEVPSFOVYRAPFGNFTKTLTACMSTVGLLQAASPSRKM--L 600
Qy 962 TPSSC-----CGLYIS--GP 974
Db 601 TTLACDARKGKMFDSFYVATKEOQHTKLYFAQSHKPPHGRVCMAPPGCLSSSDP 660
Qy 975 NKDKCPSTVNSLNCJKNCMSITMG-----SVRPSVEQPHKYLPMF---LN- 1017
Db 661 TKGFF---FVPEKVPKALSLATFGNPCVNTGCKPAVRPLVDGAMFVFGIIGVR 717
Qy 1018 ---DRENIKCPKGLAAYSTSVNLTS-----GQVL 1045
Db 718 TRQVNNHVGDP--GSVYDSSFDLFKEIGHLCHLCKAIAANSELVKPGAQCLPSGYSI 775
Qy 1046 ASRFMAVHKPLKNSD-----YTEALRAARELANITADLRKVGTDPAPFV--P 1094
Db 776 SFFLOMLHPECKELPEPNLLPGQLSHGAVGREGVQWISMAFESTTYKGS--SFQYSD 834
Qy 1095 YITTVFYEQYLTLPEG-----LFM-----LSLCL--VP 1122
Db 835 YLRMESFLQOQLALPEGSVLRGFGTCEHMKQIMEIYGVOSALCGVLSLICYAAYA 894
Qy 1123 TPAVSCLLGLDRLSGLLNLSTVIMLVDTVGFMALWDISYNAVSLINLVSAGMSVEFV 1182
Db 895 VFTTHLL---LLPVLISLIGIVLV--TIMVSGWEM--GAVEAISLIVGSSVDYC 947
Qy 1183 SHITSPAIS-----TKPTWLEBAKEATISGSAVFAGVAMTNLPGI----- 1224
Db 948 VHLVEGYLLAGENLPPHOAEDARTQRW--RTLEAVRHVGAIVSSALTTVIATVPLFC 1005
Qy 1225 LVTLGAKAQQLIQIFPFRNLTLTLGLHGVFLPVILSYGPDVNPALALOKRAE--E 1282
Db 1006 TIAPFAK-----FGKIVALNTGVSILYTLTVSTALGIMAPS-----SFRTRTSFLK 1053
Qy 1283 AVAAVAVVA 1290
Db 1054 ALGAVLLA 1061
RESULT 5
US-11-242-459-40
; Sequence 40, Application US/11242459
; Publication No. US20060035835A1
; GENERAL INFORMATION:
; APPLICANT: TANIYAMA, Yoshio
; APPLICANT: KITA, Shunobu
; APPLICANT: SATOMI, Tomoko Komiyama
; TITLE OF INVENTION: Novel Protein, Process for Producing The Same And Use Thereof
; FILE REFERENCE: 2703USOP
; CURRENT APPLICATION NUMBER: US/11/242,459
; PRIOR FILING DATE: 2005-10-03
; PRIOR APPLICATION NUMBER: US/10/239,316
; PRIOR FILING DATE: 2002-09-19
; PRIOR APPLICATION NUMBER: PCT/JP01/02279
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: JP2000-088595
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 59
; SEQ ID NO 40
; LENGTH: 1392
; TYPE: PRT
; ORGANISM: Human
US-11-242-459-40
Query Match 2.8%; Score 195.5; DB 7; Length 1392;
Best Local Similarity 19.5%; Pred. No. 7.9e-07;
Matches 200; Conservative 129; Mismatches 320; Indels 379; Gaps 47;
Qy 576 LNNYPAGDRLAQAALMEBAFLBEMRAFORRNAQMFQVTTAERSLED----- 623

Db 405 LPNYSVDR-----WEB-----ORAKFQSFVTVYAMLMKQSTSKVQLYGGTD 449
Qy 624 ---SINRTTAEED--LPFATSYIVIFLYISLALGSYSWSRWVDSKATLGLG--VAVV 676
Db 450 LFDEYVRTFPNNDMLAFISSICIALVYIILTSCVSPISFGI-----ASIGSLCLVALF 504
Qy 677 LGAVNANGFFSYLGIRSSVILQVPPVLUSVGDNIFIFVLEYQRLP-R-PGEPREXH 735
Db 505 LHVHVEGI---QYGLILNG-----VAAFIIVGIGVDVFEVINTYRQATHEDDPQLRMH 556
Qy 736 IGRALGRAVPSMLTCSLEAICFPLGALTPMPAVRTALTSGLAVI---LDELQMSAFV 792
Db 557 TVQTNK--ATFPTSITTAAYANVFSQIPAVHDFGLFMSLIYSCCLAVLTMPLAL 613
Qy 793 ALLSDSKROEASRLDVC--CCVKQDELPPQO-----FLHWI-----TRGYVL 850
Db 614 GLMSLYLAPLESQCTSGHONCSRTSLHPFDVDVATPEQYGSPPAQCPPIYLDODIFLL 673
Qy 824 -----GEGLLGFQKAYAP-----FLHWI-----TRGYVL 850
Db 674 EYEEBPVSLBELGDSLVSVSPBGLQPASNTSGRHLIVQLQELHWMVMSAVKSRWIV 733
Qy 851 LFLALFGVSL-----YSM-----CHISVGLDOE- 874
Db 734 GLFVSIILSLVFAARLPASRAPLFRPDNTIQLVLKTNLSAEGISCTTCGSLPQK 793
Qy 875 -----LALFK-----DSYLLDY--FLELNRYPEVGAPVYFVTTLGYNF 910
Db 794 PHSLONNIRTSLEKRRGSGVPMASRPATLQDPCTVYISKYKSGHRA--VYRLSLNA 851
Qy 911 SSEAGMNAICSSAG--CNNF-----SFTOKIQVATEPPEOSYLAIPASWVDDFIML 961
Db 852 SLPAWQAVSPGDGEVPSFOVYRAPFGNFTKTLTACMSTVGLLQAASPSRKM--L 905
Qy 962 TPSSC-----CGLYIS--GP 974
Db 906 TTLACDARKGKMFDSFYVATKEOQHTKLYFAQSHKPPHGRVCMAPPGCLSSSDP 965
Qy 975 NKDKCPSTVNSLNCJKNCMSITMG-----SVRPSVEQPHKYLPMF---LN- 1017
Db 966 TKGFF---FVPEKVPKALSLATFGNPCVNTGCKPAVRPLVDGAMFVFGIIGVR 1022
Qy 1018 ---DRENIKCPKGLAAYSTSVNLTS-----GQVL 1045
Db 1023 TRQVNNHVGDP--GSVYDSSFDLFKEIGHLCHLCKAIAANSELVKPGAQCLPSGYSI 1080
Qy 1046 ASRFMAVHKPLKNSD-----YTEALRAARELANITADLRKVGTDPAPFV--P 1094
Db 1081 SFFLOMLHPECKELPEPNLLPGQLSHGAVGREGVQWISMAFESTTYKGS--SFQYSD 1139
Qy 1095 YITTVFYEQYLTLPEG-----LFM-----LSLCL--VP 1122
Db 1140 YLRMESFLQOQLALPEGSVLRGFGTCEHMKQIMEIYGVOSALCGVLSLICYAAYA 1199
Qy 1123 TPAVSCLLGLDRLSGLLNLSTVIMLVDTVGFMALWDISYNAVSLINLVSAGMSVEFV 1182
Db 1200 VFTTHLL---LLPVLISLIGIVLV--TIMVSGWEM--GAVEAISLIVGSSVDYC 1252
Qy 1183 SHITSPAIS-----TKPTWLEBAKEATISGSAVFAGVAMTNLPGI----- 1224
Db 1253 VHLVEGYLLAGENLPPHOAEDARTQRW--RTLEAVRHVGAIVSSALTTVIATVPLFC 1310
Qy 1225 LVTLGAKAQQLIQIFPFRNLTLTLGLHGVFLPVILSYGPDVNPALALOKRAE--E 1282
Db 1311 TIAPFAK-----FGKIVALNTGVSILYTLTVSTALGIMAPS-----SFRTRTSFLK 1358
Qy 1283 AVAAVAVVA 1290
Db 1359 ALGAVLLA 1366
RESULT 6

Matches	69;	Conservative	41;	Mismatches	106;	Indels	56;	Gaps	12
Qy	576	LNTYPADGPRLIAQAKLWEBAFLXKRAFORMAGNFQVFTTAERSLED-----							623
Db	100	LPNYSYVDR-----WEE-----QRAKFOGFVTTYVAMLAKOSTSKVLYGGTD							144
Qy	624	-----EINRTTAEED--LPFATSYIYIFLYISLALSYSGWSRWVDSKATLGLG--VAVV							676
Db	145	LPDEVERTTFNNMDLAFISSCIAALYILTSCVFLSFQI-----ASIGLGLVALF							199
Qy	677	LGAVMAMAGFFSYLIGIRSSLVILQVPPVLVSVGDNIFFIPLYEQRLPR--RPGEPREVV							735
Db	200	LYHVHVFGI---QYGLILNG-----VAAAVIYIGIDVDVFEFNTYRQATHEEDPOLRMIH							251
Qy	736	IGRALGRVAPMLTCLSLSEALCFEGALTEPMVAVTFTALSGLAVI---LDFILQMSAFV							792
Db	252	TVQTRAGK---ATFFSLTTLTAAYANAVNSQIPAVHDFGLFMSLIYSCMLAVLYMPAAL							308
Qy	793	ALLISDSRKQBASRLDVC--CCVRKQQLPPPG							822
Db	309	GLMSLYLAPLESSCQTSCHQNSRRTTSLHPFG							340

```

RESULT 10
US-10-392-234A-18
; Sequence 18, Application US/10392234A
; Publication No. US20050255538A1
; GENERAL INFORMATION:
; APPLICANT: Pharmacia and Upjohn Corporation
; APPLICANT: Buxser, Steven
; APPLICANT: Poole, Keith
; APPLICANT: Decker, Douglas
; APPLICANT: Xiazanhi Li
; TITLE OF INVENTION: Method for Screening for acrAB Transporter Family Inhibitors
; FILE REFERENCE: 6206
; CURRENT APPLICATION NUMBER: US/10/392,234A
; CURRENT FILING DATE: 2003-03-17
; PRIOR APPLICATION NUMBER: US 60/364,935
; PRIOR FILING DATE: 2002-03-15
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 18
; LENGTH: 1048
; TYPE: PRT
; ORGANISM: Salmonella typhimurium
US-10-392-234A-18

```

Query Match	2.0%;	Score 141.5;	DB 6;	Length 1048;
Best Local Similarity	18.1%;	Pred. No. 0.0097;		
Matches 145;	Conservative 110;	Mismatches 243;	Indels 303;	Gaps 32

```

QY 624 SINRTTAEPLFIANSYIYIFLYISLADSYSMSXWMDSKTL--GGGVAWVG--A 679
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 339 EVVKTLVE-----AIIYELWMIYELONF-----RATLPTTAVPVLLGTRA 381
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 680 VMAANGFESYLGIRSSLVLQVPPFLVLSVG--ADNIFIVLEYOHLPRRPGEPREWHIG 737
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 382 VLAARGF-----SINTLTWEG--WVLAIGLLVDALIVVENVEKVMTEBGLPPREATR 432
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 738 PALGR-----VAPSMILCSLSAICFFLFCALTPMPAVRFPALTSGIAVILDFLLQMSAFY 792
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 433 KSMGCIQCALVGIAMVLSAVFIPLMAFFGSGSTGAI--YRQFSITIVSAMALSVLVALILTP 490
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 793 ALLSIDSQRQESRLDVCCCVRPQELPRPGQSGILLGFGQKYAVFLLHMI----- 844
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 491 ALCA-----TMLKPVAKGDHGEKKGFEGFMFNRIEDKSTHHYDVSQNIILR 536
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 845 TRGVLLLFALF-----GVSLYSMCHISVGLDDELAPKDSYLL 884
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 537 STGRLLILLYLIVGMAYLFVRLPSSFLPDEDDQGVFL--IMVQLPAGATGEROKXLDVYI 595
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 885 DYFL-----FLNR----- 892
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db      556 DYILNKEANVESFAVAVNGPGRAGGQNTGIAFVSLKMDADRGKKNVEALTQARATAF 655
Qy      893 -----YEVGAPVYFVTTLTGYNESSEAGMNAICSSAGCNFSPFO----- 932
Db      656 SQIKDAMFAENLPAIVELGTATGDFE-----LIDQAGIGHKLTQARNQLEGEVAK 708
Qy      933 -----KIQVATEPREOSYLAIPAS-----SWVDFDWTLP 963
Db      709 YPDLLVYGRPNGLBTPQFKIDIDOKKQALGVSLSDINTTLTGAMGSSYVDFID--R 765
Qy      964 SSCRLYISGENKDKFCPCSTVNSLNCJKNCSITWGSVRSVEQFKHYLPWFELNDRPIK 1022
Db      766 GRVKKVYVMSAKYRMLPDDIND-----WYRGSDDQM 798
Qy      1024 CPKGLAAYR-----SVNLTSDQVLAARFMAHYKPLKNSQDYTEALRARBELAANT 107
Db      799 VP--FSAFSSSRMEYGSPLREYNGALPSMELLGOALGKST--GEANAMMEBLASKUP 852
Qy      1078 ADLRKVPETDPAFEVFPYITNVFEQYLT-----LPEGLFMLSCL----- 1124
Db      853 SGI-----GYDWTGMSYQERLSGQAPALVAISLIYVFLCLALYESWSIPF 899
Qy      1121 -----VPPFAVSCLLIGDLASGLLNLSTVMILVDYTGFMALMDISTYAVSLIN----- 1177
Db      900 SVMLVPLVPGVIGALLA--TFRGLTNDVYFOVGLTTLTGLSA--KNAILIVFAPDL 952
Qy      1171 -----LVSAVGSVEFVSHITRSPASTPTMLBRAKEATISMGSAVFAVAMT 121
Db      953 MDKSGKGLVEAMLEAVRMLRPLIMLTSLAFMLGVNP-----LVISSG--AGSGAQ 1004
Qy      1220 NLPGLIV-GLAKAOIPIF 1239
Db      1001 NAVGTGVIGMVTATVLAIFP 1021

```

```

RESULT 11
US-10-392-234A-16
; Sequence 16, Application US/10392223A
; Publication No. US2005025538A1
; GENERAL INFORMATION:
; APPLICANT: Pharmacia and Upjohn Corporation
; APPLICANT: Buxser, Steven
; APPLICANT: Poole, Keith
; APPLICANT: Decker, Douglas
; APPLICANT: Kiazhi, Li
; TITLE OF INVENTION: Method for Screening for Acarab Transporter Family Inhibitors
; FILE REFERENCE: 6206
; CURRENT APPLICATION NUMBER: US/10/392,234A
; CURRENT FILING DATE: 2003-03-17
; PRIOR APPLICATION NUMBER: US 60/364,935
; PRIOR FILING DATE: 2002-03-15
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16
; LENGTH: 1046
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-10-392-234A-16

```

Query Match 2.0%; Score 140.5; DB 6; Length 1046;

Best Local Similarity 20.5%; Fred. NO. 0.042;
Matches 187; Conservative 126; Mismatches 305; Indels 293; Gaps 49

506 NQTLMGQT--SQVDWKDHFYCANAPLTFKDGTAALSCMADYGAPVFPFLAIGGYKGD 563

Db 231 NATIGKTRLQTAQFENILKVN-----PDGSQVRLKDVADV-----LG-----GQD 274

QY 564 YSEAEALIMTFSLNYP-----GDPRLAQAKLWEE--AFLEEMRAFQRMAGMFQ 612

Db 275 YS-----INAQFNGSPASGIAIKLATGANALDTAKAIRQTIANLEPFMPQGMKVVPYD 328

QY 613 VTFTAERSLEDEINRTTAEDLPFATSYIVIFLYISLALGSSWSRVMVDSKATL--GL 670

```

Db      329 TTPVVSASH-EVVKTLGE-----ALLVFLVMTLFLONF-----RATLIPIT 370
Qy      671 GGVAVVIG--AVMAAGFPSYLGIRSSLVLQVPPVLVSVG--ADNIFIVLEAYRLPR 726
      371 AVPVVLIGTFGVLAARGF-----SINTLTMEG--MVLALIGLVDAIVVENVERVMA 421
Qy      727 REG-EREV---HIGALGR-VAPSMILCSLSAICFPLGALTPMAVTPALTSGLAVI 781
      422 EREGLSREARAKSMGOIGALVGIAMVLSVFLPMAFFGGSTGVI--YRQFSITIVSAMA 479
Qy      782 LDFLLQMSAFVALLSLDSKRQGEASRLDVCC---VPRQELPPRQGEGLLGFPOKAYAP 838
      480 L-----SVTVALL-----LTPALCATMLKPIEGDGHGKGFGEFNMKFLS 522
Qy      839 FLIMWITRGV-----LILFLALFGVSLYSMCHISVGL-----DQ-----ELALPKD 880
      523 -TTHGVERGVASTLKKRAPYLLLVIVVAGMIMFRIPIPAFLPBDQGLVFAQVQTPPG 581
Qy      881 S-----YLDYFLFLNRYFEVGADEVYFTTLGINPSSB-----913
      582 SSARFTQVAVDSMRREYLEK-----ESSSVSVFTVGFNFAGRGQSSGMAFTMLKP 633
Qy      914 -----AGMAI---CSSAGCNNSFTOKIOYATEFBQSYLALPASSWDDFD-----959
      634 WEERPGSEMSVFEELARAKQHFPSPFXDAMVFA--FAPPSVLELGNATGFDLQDQAGVG 691
Qy      960 ---WLTSSCCCLYISGPNKDFPCSTVA-----SLNCLKKCMST 996
      692 HEVTLQARKKFTMLAQNAPALQVRPNQMSDEPOYKLEIDDEKASALGVSGLADINSTVSI 751
Qy      997 TWGSV-----RPSVEQFHXYLPWFLNDRPN-----IKCPKGLA---AYST 1034
      752 AMGSSVYNDPDIRGRVYR-----YLQGRPDARMPDDLSKMYVNRDKEMVPFNAFAT 805
Qy      1035 -----SVNLTSGQVLASFMAYHKPKRSQDYTEALRAARLANITDIAKVGTDPE 1088
      806 GKREYSPKLERYNGVPAHEILGEPAIGLSSGD---AMAAVEB---YVQOLPKGVG---855
Qy      1089 AEFVFPYTTITNVFEQYLT-----LLEBGLFMLSIC-----LVPTFAVS 1127
      856 -----YSWTGLSYERLSGSOQAPALYALSLVVFCLALYESWSPFVMTLVPLGVI 909
Db      1128 CLLGLDLSGGLNLSTIWMILVDIVGFMALMDISYNAVSLNL-----V 1172
      910 GALLATSMR-GLSNDVFFQVGLTTIGLSA-----KNAILIVEFAKELHEQKGIYEAL 963
Qy      1173 SAVGMSVEFVSHTTRSPALSTKPTWLBRAKEATISMGSAVFAVAMTNLPGLIVL-GLAK 1231
      964 EACRMKRLRPIVMTSLAFILGVDP-----LAIISG---AGSGQHAIGTGVIGGMT 1011
Db      1232 AOLIOIFFPRL 1242
Qy      1012 ATVLAIFWVPL 1022

```

```

RESULT 12
US-10-467-657-2648
; Sequence 2648, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIANT Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04

```

```

; SEQ ID NO 2648
; LENGTH: 1067
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-2648

Query Match      1.9%; Score 134; DB 6; Length 1067;
Best Local Similarity 19.4%; Pred. No. 0.039;
Matches 185; Conservative 139; Mismatches 354; Indels 274; Gaps 44;

Qy      452 RHLQVW-SPEAQR--NISLQDICYPALNPDNSTLYVCCINSLLQYQNNRTLLLTANOT 508
      180 RMRIVWDPRKQIQNTVLSPADVSA-LSAQNTQISGSGSLPAV--RCGT---VTRATV 233
Qy      509 LMGQTSQVDMKQHFLYCANAPLTFKQGTALALSQMDYGAVPVFPALIGYKQDYSEAE 568
      234 AQQLGTAEFQGVILARNT-----DGSNIYLKDVAKVGLGM-----EDYSSST 277
Qy      569 ALI-----MTSLNNYPAGDPRLOAKLWEAFLEEMRAFORRMAGMFOVTTAERSLE 622
      278 RLNGVNTTGAAVWLSN--SGN-AMATAKAVKERLAVLEKYPQGMG--WKTPTYDSKFEV 332
Qy      623 DEINRTAADLPIFATSYVIEFLYISLALGSYSMSGRVWDSKATIGLGSVAIVLGAVMA 682
      333 ISEKVIHTLIEAMVLVFWMTVLEONT--RYTLIFTVIPISL---LGSFA-----379
Qy      683 AMGFPSYLGIRSSLVLQVPPVLVSVDGADNIFIIVLEAYQRLPRRQGEPREVHIGRALGR 742
      380 ---FISYMGH-SINVLTMFAMILVIGIVDDALVVENNERIWAAGGLPKPKATKKAMGQ 435
Qy      743 VAPSMU-----LCSLSAICFPLGALTMPMPAVRTPLTSGLAVIDLPLQMSAFVALLSL 797
      436 ISGAVIGITAVLVISVFPVLPMSGAGNI--YKQFPLTMASST-----AFSAFLAL--484
Qy      798 DSKRQASRLDVCCCKPQELPPQGGEGI--LLGFPOKAYAP-----LHMWIT 845
      485 -----LTPALCATMLKTTIPKGHHEKGFGEFNMKFKDSWTHGYBQVAKVLRKT 536
Qy      846 RGVALLFLALFGVSLYSMCHISVGLDOELAPKDSYLLDYPLNRYFEVGAVPYFVTT 905
      537 RMNVVYIGLAVGV-----LFLMR-----LPTSGLPT 563
Qy      906 LQYNSFSEAGMNAICSSAGC---NNFSFTOKIOYATEPPE-OSYLAIPASSVVD---966
      564 EDQGFV-----MVSQAPAGATKERTDATLAQVTLAKSIPELENIITVSGSFSSGSGQNM 619
Qy      957 -----FIDW--LTPSSCCRLYIS-----QPNQKPCPSVYN-----S 986
      620 AMGFALIKQWNERTAGSDAIVAVAGKLTGMWMTLKDGFQISVPPPIELNGSGSLSTN 679
Qy      987 LNCCLKMSTITWGSVAPSVQFHKYLPWF-----LNDRENIKCPKGLAAVSTSVN 1037
      680 LODRNNITGALLAKKNELIQKRASGLFDPSTVRAGGLEDSQKIDINRAAAAAGIS 739
Qy      1038 LPSDQVLA-SRF-MAYHKPLKNSQDYTEALRAARELANITDILRK-----VP--1084
      740 FADIRTAALASLSSVSDPFPNGRLQRYVWQEDBARQPADILMLTVPNKSGVAVPLS 799
Qy      1085 -----GTD-----PAEVFPYTTITNVFEQYLT-----1106
      800 TIATVSWENGTEQSVFENGVPMSKLSASPATGVSTQQAAMAAQVMDLGGGYSPFMGQ 859
Qy      1107 -----TILBGLFMLSICLV-----PTFAVSCLLGL-----DLR 1136
      860 SREKAGSGOTLLIYGLAIVAVLVLAALYESMSTILAVILVPLGLIGAAAGVTRNLPE 919
Qy      1137 SGLL-NILSITWMLVDIVGFMALMDIS-YNAVSLNLV-----SAVMSVE-----F 1181
      920 EGLGVSVPFANDIYQVGFVTVMGSAKNAIILIFPAKDLOQKGSAAVEAALAEARLRF 979
Qy      1182 VSHITRSFA--ISTKPTWLBRAKEATI--SMGSAVAGVAMTNLPGLIVLGL 1229
      980 RPIIMTSFAFILGVPLVPLIYAAGSSASQRAIGTIVWGMVLVGTLLSVFLVPL 1031

```


THIS PAGE BLANK (w8pt0)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using SW model

Run on: March 22, 2006, 23:06:57 ; Search time 188 Seconds
(without alignments)
2960.365 Million cell updates/sec

Title: US-10-736-769-4

Perfect score: 6909
Sequence: 1 MAEAGIRGWLWALLRLAQLA.....GSIKAGAIISNPLPNNRGPF 1332

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA_Main:*
1: /cgn2_6/prodata/1/pubppaa/US07_PUBCOMB.pep:*
2: /cgn2_6/prodata/1/pubppaa/US08_PUBCOMB.pep:*
3: /cgn2_6/prodata/1/pubppaa/US09_PUBCOMB.pep:*
4: /cgn2_6/prodata/1/pubppaa/US10A_PUBCOMB.pep:*
5: /cgn2_6/prodata/1/pubppaa/US10B_PUBCOMB.pep:*
6: /cgn2_6/prodata/1/pubppaa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6909	100.0	1332	4 US-10-621-758A-4	Sequence 4, Appl1
2	6909	100.0	1332	4 US-10-663-208A-4	Sequence 4, Appl1
3	6909	100.0	1332	4 US-10-646-301A-4	Sequence 4, Appl1
4	6909	100.0	1332	4 US-10-736-769-4	Sequence 4, Appl1
5	6896	99.8	1332	4 US-10-239-316-9	Sequence 9, Appl1
6	6872.5	99.5	1359	4 US-10-621-758A-4	Sequence 44, Appl1
7	6872.5	99.5	1359	4 US-10-663-208A-4	Sequence 44, Appl1
8	6872.5	99.5	1359	4 US-10-646-301A-4	Sequence 44, Appl1
9	6872.5	99.5	1359	4 US-10-736-769-4	Sequence 44, Appl1
10	6536	94.6	1334	5 US-10-450-763-53052	Sequence 53052, A
11	5421.5	78.5	1331	4 US-10-621-758A-2	Sequence 2, Appl1
12	5421.5	78.5	1331	4 US-10-663-208A-2	Sequence 2, Appl1
13	5421.5	78.5	1331	4 US-10-646-301A-2	Sequence 2, Appl1
14	5421.5	78.5	1331	4 US-10-736-769-2	Sequence 2, Appl1
15	5407	78.3	1333	4 US-10-621-758A-12	Sequence 12, Appl1
16	5407	78.3	1333	4 US-10-663-208A-12	Sequence 12, Appl1
17	5407	78.3	1333	4 US-10-646-301A-12	Sequence 12, Appl1
18	5407	78.3	1333	4 US-10-736-769-12	Sequence 12, Appl1
19	4466	64.6	982	5 US-10-450-763-53050	Sequence 53050, A
20	2402.5	34.8	1278	4 US-10-208-731-2	Sequence 2, Appl1
21	2402.5	34.8	1278	4 US-10-741-601-530	Sequence 530, App
22	2402.5	34.8	1278	5 US-10-741-600-1542	Sequence 1542, App
23	2402.5	34.8	1278	5 US-10-756-149-4924	Sequence 4924, App
24	2385	32.7	1319	4 US-10-208-731-4	Sequence 4, Appl1
25	2262	32.7	1287	6 US-11-097-143-12003	Sequence 12003, A
26	1864.5	27.0	1223	6 US-11-097-143-2679	Sequence 2679, App
27	1329	19.2	1170	4 US-10-208-731-6	Sequence 6, Appl1

28	1078	15.6	1296	4 US-10-208-731-9	Sequence 9, Appl1
29	942	13.6	194	4 US-10-239-316-8	Sequence 8, Appl1
30	900.5	13.0	492	4 US-10-424-599-211862	Sequence 211862, Appl1
31	859.5	12.4	541	4 US-10-424-599-189288	Sequence 189288, Appl1
32	651.5	9.4	1447	2 US-08-954-701A-19	Sequence 19, Appl1
33	651.5	9.4	1447	3 US-09-898-533-5	Sequence 5, Appl1
34	651.5	9.4	1447	3 US-09-754-032-19	Sequence 19, Appl1
35	651.5	9.4	1447	4 US-10-421-446-19	Sequence 19, Appl1
36	651.5	9.4	1447	4 US-10-791-844-6	Sequence 6, Appl1
37	646.5	9.4	1434	2 US-08-954-701A-10	Sequence 10, Appl1
38	646.5	9.4	1434	4 US-09-754-032-10	Sequence 10, Appl1
39	646.5	9.4	1434	4 US-10-421-446-10	Sequence 10, Appl1
40	628	9.1	1296	4 US-10-302-279-60	Sequence 60, Appl1
41	598.5	8.7	933	4 US-10-415-934-3	Sequence 3, Appl1
42	597	8.6	821	5 US-10-890-776A-4805	Sequence 4805, App
43	588	8.5	954	4 US-10-060-756A-3	Sequence 3, Appl1
44	588	8.5	954	5 US-10-890-776A-3	Sequence 3, Appl1
45	584	8.5	1203	3 US-09-990-046-2	Sequence 2, Appl1

ALIGNMENTS

RESULT 1									
US-10-621-758A-4									
; Sequence 4, Application US/10621758A									
; Publication No. US20040093629A1									
; GENERAL INFORMATION:									
; APPLICANT: Altmann, Scott W									
; APPLICANT: Wang, Luquan									
; APPLICANT: Graziano, Michael									
; APPLICANT: Margolio, Nick									
; TITLE OF INVENTION: NPC1L1 (NPC3) AND METHODS OF USE THEREOF									
; FILE REFERENCE: JB01603-K-US									
; CURRENT APPLICATION NUMBER: US/10/621,758A									
; CURRENT FILING DATE: 2003-07-17									
; PRIOR APPLICATION NUMBER: 60/397,442									
; PRIOR FILING DATE: 2002-07-19									
; NUMBER OF SEQ ID NOS: 50									
; SOFTWARE: PatentIn version 3.1									
; SEQ ID NO: 4									
; LENGTH: 1332									
; TYPE: PRT									
; ORGANISM: Homo sapiens									
US-10-621-758A-4									
Query Match 100.0%; Score 6909; DB 4; Length 1332;									
Best Local Similarity 100.0%; Pred. No. 0;									
Matches 1332; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	MAEAGIRGWLWALLRLAQLAQLA.....GSIKAGAIISNPLPNNRGPF	60						
DB	1	MAEAGIRGWLWALLRLAQLAQLA.....GSIKAGAIISNPLPNNRGPF	60						
QY	61	TPARKITGHTLILQKICPLRYTPNTQACSAKQVLSIASITKALLTRCPACSDNF	120						
DB	61	TPARKITGHTLILQKICPLRYTPNTQACSAKQVLSIASITKALLTRCPACSDNF	120						
QY	121	VNIHCNNTGSPNOSLINTRVVQLGAGQLPAVVAEATFQHSFAEOSYDSCRRVPPAA	180						
DB	121	VNIHCNNTGSPNOSLINTRVVQLGAGQLPAVVAEATFQHSFAEOSYDSCRRVPPAA	180						
QY	121	VNIHCNNTGSPNOSLINTRVVQLGAGQLPAVVAEATFQHSFAEOSYDSCRRVPPAA	180						
DB	121	VNIHCNNTGSPNOSLINTRVVQLGAGQLPAVVAEATFQHSFAEOSYDSCRRVPPAA	180						
QY	181	ATLAVGTCGVVSALCNARWLNFGDGTNGALPDIPTFHLEPGQAVSGIQPLNEGV	240						
DB	181	ATLAVGTCGVVSALCNARWLNFGDGTNGALPDIPTFHLEPGQAVSGIQPLNEGV	240						
QY	241	ARCNEQSDVATVATCSQDCAASCPAATAPQALDSTFYLGMPGSLVLIILCSVFAVVTI	300						
DB	241	ARCNEQSDVATVATCSQDCAASCPAATAPQALDSTFYLGMPGSLVLIILCSVFAVVTI	300						
QY	301	LIVGFRVAPARDKSKVNDPKKGTSLSDKLSFSHTILLGPFQGWGTWVSWPLTIVLSV	360						
DB	301	LIVGFRVAPARDKSKVNDPKKGTSLSDKLSFSHTILLGPFQGWGTWVSWPLTIVLSV	360						

QY 361 IIVVAAAGLVFTELTITDPEVEMSAFNSQARSEKAFHDOHFGFPRFTNOVILTAENRSSY 420
DB 361 IIVVAAAGLVFTELTITDPEVEMSAFNSQARSEKAFHDOHFGFPRFTNOVILTAENRSSY 420
QY 421 RYDSLLGPKNFSGLIDDLLELLELORLRLHLYQWSPBAORNISLODI CYAPLNPDMT 480
DB 421 RYDSLLGPKNFSGLIDDLLELLELORLRLHLYQWSPBAORNISLODI CYAPLNPDMT 480
QY 481 SYDDCINSLLQYFQNNRRLTLTLTANQTMGOTSQVDMKDHFLYCANAPLTFKDGATALL 540
DB 481 SYDDCINSLLQYFQNNRRLTLTLTANQTMGOTSQVDMKDHFLYCANAPLTFKDGATALL 540
QY 541 SCMDYGAAPVFPFLAIGYKKGKDYSEAEALIMTFSLNNYPADGPRLAOKLMEAEFLBEM 600
DB 541 SCMDYGAAPVFPFLAIGYKKGKDYSEAEALIMTFSLNNYPADGPRLAOKLMEAEFLBEM 600
QY 601 RAFORMAMGFQVTFPFAERSLDEINRTTAEDLPITATSYIVIFLYISIALGYSYSSMSRV 660
DB 601 RAFORMAMGFQVTFPFAERSLDEINRTTAEDLPITATSYIVIFLYISIALGYSYSSMSRV 660
QY 661 MYSKATTLGLGVAAYVLGAVMAMGFSSYLGRSSLVIIQVVPFLVLSGADNIFIFVLE 720
DB 661 MYSKATTLGLGVAAYVLGAVMAMGFSSYLGRSSLVIIQVVPFLVLSGADNIFIFVLE 720
QY 721 YORLPERRPGEPRVHIGRALGRVAPSMILCSLSEALICFPLGALITMPAVRTFALISGLAV 780
DB 721 YORLPERRPGEPRVHIGRALGRVAPSMILCSLSEALICFPLGALITMPAVRTFALISGLAV 780
QY 781 IILDFLOMSAFVALLSLDSKROBASRLDVCCCVKPELPPGGGBGLLGFCKAAPL 840
DB 781 IILDFLOMSAFVALLSLDSKROBASRLDVCCCVKPELPPGGGBGLLGFCKAAPL 840
QY 841 LHMITRGVLLFLFALFGVSLYSMCHI SVGLDOELALPKDSYLLDYFLFLNRYFEVAVY 900
DB 841 LHMITRGVLLFLFALFGVSLYSMCHI SVGLDOELALPKDSYLLDYFLFLNRYFEVAVY 900
QY 901 YFVTTTLYGNFSSSEAGNNAICSSAGCANNFSTQKI QYATEPPEQSYLAIPASSVWDPIWM 960
DB 901 YFVTTTLYGNFSSSEAGNNAICSSAGCANNFSTQKI QYATEPPEQSYLAIPASSVWDPIWM 960
QY 961 LTPSSCCRLYISGPNKDXCPSTVNSLNLCKNCKMSITMGSVSPSVOPHKYLPWFLNDRP 1020
DB 961 LTPSSCCRLYISGPNKDXCPSTVNSLNLCKNCKMSITMGSVSPSVOPHKYLPWFLNDRP 1020
QY 1021 NIKCPKGLAAAYSTSVNLSDQVLA SRFMAHYHKKNSQODYTEALRAARELANITADL 1080
DB 1021 NIKCPKGLAAAYSTSVNLSDQVLA SRFMAHYHKKNSQODYTEALRAARELANITADL 1080
QY 1081 RAVPGTDPAPFVPPYTTITNVFEBOYLTILPEGLFMLSCLIVPTFAVSCLLGLDLSGILL 1140
DB 1081 RAVPGTDPAPFVPPYTTITNVFEBOYLTILPEGLFMLSCLIVPTFAVSCLLGLDLSGILL 1140
QY 1141 NLLSTMIILVDPMALMDISYNAVSLTNLVAQMSVEFVSHITRSPAISXKPLWBLR 1200
DB 1141 NLLSTMIILVDPMALMDISYNAVSLTNLVAQMSVEFVSHITRSPAISXKPLWBLR 1200
QY 1201 AKEATISMSGSAVAGVAMTNLPGLIVLGLAKAQLIOIFFERLNLITLLGLHGLVFLVY 1260
DB 1201 AKEATISMSGSAVAGVAMTNLPGLIVLGLAKAQLIOIFFERLNLITLLGLHGLVFLVY 1260
QY 1261 ILSYVGPVNPALALEQKRAEBAVAAVMVASCPNHPRSVSTADNIYVNSHFEKIGAGA 1320
DB 1261 ILSYVGPVNPALALEQKRAEBAVAAVMVASCPNHPRSVSTADNIYVNSHFEKIGAGA 1320
QY 1321 ISNPLPNNGRQF 1332
DB 1321 ISNPLPNNGRQF 1332

RESULT 2
US-10-663-208A-4
; Sequence 4, Application US/10663208A

Publication No. US20040132058A1
; GENERAL INFORMATION:
; APPLICANT: Altman, Scott W
; APPLICANT: Wang, Luquan
; APPLICANT: Graziano, Michael
; APPLICANT: Muirgo, Nick
; TITLE OF INVENTION: NPC11 (NPC3) AND METHODS OF USE THEREOF
; FILE REFERENCE: JB01603K2 US
; CURRENT APPLICATION NUMBER: US/10/663,208A
; PRIOR FILING DATE: 2003-09-16
; PRIOR APPLICATION NUMBER: 60/397,442
; PRIOR FILING DATE: 2002-07-19
; PRIOR APPLICATION NUMBER: 10/621,758
; PRIOR FILING DATE: 2003-07-17
; PRIOR APPLICATION NUMBER: 10/646,301
; PRIOR FILING DATE: 2003-08-22
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 4
; LENGTH: 1332
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-663-208A-4
Query Match 100.0%; Score 6909; DB 4; Length 1332;
Best Local Similarity 100.0%; Freq. No. 0;
Matches 1332; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAEAGLRGWLWALLRLRLAQSPPYTTIHQPGYCAFYDECKNPELSGSLMTLSNVSCLSN 60
DB 1 MAEAGLRGWLWALLRLRLAQSPPYTTIHQPGYCAFYDECKNPELSGSLMTLSNVSCLSN 60
QY 61 TPARKITGDHLLLOKICRPLTYGNTQACCSAKOLVSIASLSTTKALLTRCPACSDNF 120
DB 61 TPARKITGDHLLLOKICRPLTYGNTQACCSAKOLVSIASLSTTKALLTRCPACSDNF 120
QY 121 VNLHCNNTCSPNQSLFINTRTRVAOLGAGQLPVAVVEAFYQHSFPAEQSDSCSRVRVPA 180
DB 121 VNLHCNNTCSPNQSLFINTRTRVAOLGAGQLPVAVVEAFYQHSFPAEQSDSCSRVRVPA 180
QY 181 ATLAVGTMCGVYGSALCNARWLNFQDGTGNGLAPLDIRFHLLPEGOAVSGSIQPLNEGV 240
DB 181 ATLAVGTMCGVYGSALCNARWLNFQDGTGNGLAPLDIRFHLLPEGOAVSGSIQPLNEGV 240
QY 241 ARCNSQGDVATCSQODCAASCPAIARPOLDSFFYLGOMPGSLVLIILCSVPAVVTI 300
DB 241 ARCNSQGDVATCSQODCAASCPAIARPOLDSFFYLGOMPGSLVLIILCSVPAVVTI 300
QY 301 LLVGRVVAAPARSKRWVPKKTSLSDKLSFSTHTLLGQFPGQKGTWASWPLTILVLSV 360
DB 301 LLVGRVVAAPARSKRWVPKKTSLSDKLSFSTHTLLGQFPGQKGTWASWPLTILVLSV 360
QY 361 IIVVAAAGLVFTELTITDPEVEMSAFNSQARSEKAFHDOHFGFPRFTNOVILTAENRSSY 420
DB 361 IIVVAAAGLVFTELTITDPEVEMSAFNSQARSEKAFHDOHFGFPRFTNOVILTAENRSSY 420
QY 421 RYDSLLGPKNFSGLIDDLLELLELORLRLHLYQWSPBAORNISLODI CYAPLNPDMT 480
DB 421 RYDSLLGPKNFSGLIDDLLELLELORLRLHLYQWSPBAORNISLODI CYAPLNPDMT 480
QY 481 SYDDCINSLLQYFQNNRRLTLTLTANQTMGOTSQVDMKDHFLYCANAPLTFKDGATALL 540
DB 481 SYDDCINSLLQYFQNNRRLTLTLTANQTMGOTSQVDMKDHFLYCANAPLTFKDGATALL 540
QY 541 SCMDYGAAPVFPFLAIGYKKGKDYSEAEALIMTFSLNNYPADGPRLAOKLMEAEFLBEM 600
DB 541 SCMDYGAAPVFPFLAIGYKKGKDYSEAEALIMTFSLNNYPADGPRLAOKLMEAEFLBEM 600
QY 601 RAFORMAMGFQVTFPFAERSLDEINRTTAEDLPITATSYIVIFLYISIALGYSYSSMSRV 660
DB 601 RAFORMAMGFQVTFPFAERSLDEINRTTAEDLPITATSYIVIFLYISIALGYSYSSMSRV 660
QY 661 MYSKATTLGLGVAAYVLGAVMAMGFSSYLGRSSLVIIQVVPFLVLSGADNIFIFVLE 720

```

|||||
Db 661 MVDKATLIGGVAVVLGAVMAAMGFSSYLGRSSLVIIQVVPFLVLSVGADNIFIFVLE 720
Qy 721 YQRLPRRPGPREVHIGRALGRVAPSMULCSLSEALICFPLGALTTPMPAVTTPALTSLAV 780
Db 721 YQRLPRRPGPREVHIGRALGRVAPSMULCSLSEALICFPLGALTTPMPAVTTPALTSLAV 780
Qy 781 IIDFLQMSAFVALLSLDSKROBASRLDYCCCVKPOELPPPGGEGILLGFQKAVAPFL 840
Db 781 IIDFLQMSAFVALLSLDSKROBASRLDYCCCVKPOELPPPGGEGILLGFQKAVAPFL 840
Qy 841 LHMITGVVLLFLALFGVSLYSMCHISVGLDOELAPKDSYLLDYFLPLNRYFEVGA 900
Db 841 LHMITGVVLLFLALFGVSLYSMCHISVGLDOELAPKDSYLLDYFLPLNRYFEVGA 900
Qy 901 YFVTTTGYNFSSEAGMNAICSSAGCNNSFTQKIQVATEPPEQSYLAIPASSWVDFIDW 960
Db 901 YFVTTTGYNFSSEAGMNAICSSAGCNNSFTQKIQVATEPPEQSYLAIPASSWVDFIDW 960
Qy 961 LTPSSCCRLYISGPNKDKFCPTVNSLNCCKMSITMGSVRPSVEQFHKYLPWFLNDRP 1020
Db 961 LTPSSCCRLYISGPNKDKFCPTVNSLNCCKMSITMGSVRPSVEQFHKYLPWFLNDRP 1020
Qy 1021 NIKCPKGLAAYSTSVNLTSDGOVLASRFMAVYHKPLKNSODYTEALPAARELANITADL 1080
Db 1021 NIKCPKGLAAYSTSVNLTSDGOVLASRFMAVYHKPLKNSODYTEALPAARELANITADL 1080
Qy 1081 RKVPGTDPAPFEPYITITNVFYEQYLTILPEGLFMLSCLVPTFAVSCLLGLDLSGGL 1140
Db 1081 RKVPGTDPAPFEPYITITNVFYEQYLTILPEGLFMLSCLVPTFAVSCLLGLDLSGGL 1140
Qy 1141 NLISYVILVDYVGFMAIMDISYNAVSLINLVSAVMSVEFVSHITSPAISTRKPTWLER 1200
Db 1141 NLISYVILVDYVGFMAIMDISYNAVSLINLVSAVMSVEFVSHITSPAISTRKPTWLER 1200
Qy 1201 AKKATISMSGSAFVAGVAMTLPGLIVLGLAKAOLIOIFFRLNLITLLGLGLVFLV 1260
Db 1201 AKKATISMSGSAFVAGVAMTLPGLIVLGLAKAOLIOIFFRLNLITLLGLGLVFLV 1260
Qy 1261 ILSYVGPDPVAPALBEQKRAEBAVAVMAVASCNHPRSVSTADNITYNHSFEQSIKAGA 1320
Db 1261 ILSYVGPDPVAPALBEQKRAEBAVAVMAVASCNHPRSVSTADNITYNHSFEQSIKAGA 1320
Qy 1321 ISNFLPNNGRQF 1332
Db 1321 ISNFLPNNGRQF 1332

RESULT 3
US-10-646-301A-4
: Sequence 4, Application US/10646301A
: Publication No. US20040137467A1
: GENERAL- INFORMATION:
: APPLICANT: Altman, Scott W
: APPLICANT: Wang, Luquan
: APPLICANT: Graziano, Michael
: APPLICANT: Margolo, Nick
: TITLE OF INVENTION: NPC11 (NPC3) AND METHODS OF USE THEREOF
: FILE REFERENCE: J061603-K1-US
: CURRENT APPLICATION NUMBER: US/10/646,301A
: PRIOR FILING DATE: 2003-08-22
: PRIOR APPLICATION NUMBER: 60/397,442
: PRIOR FILING DATE: 2002-07-19
: PRIOR APPLICATION NUMBER: 10/621,758
: PRIOR FILING DATE: 2003-07-17
: NUMBER OF SEQ ID NOS: 50
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 4
: LENGTH: 1332
: TYPE: PRT
: ORGANISM: Homo sapiens
US-10-646-301A-4

```

```

Query Match 100.0%; Score 6909; DB 4; Length 1332;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1332; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAEAGLRGMLMALRLAQLQSEPTTTHQPGYCAFDEGCKNPELSGSLMTLSNVCSLN 60
Db 1 MAEAGLRGMLMALRLAQLQSEPTTTHQPGYCAFDEGCKNPELSGSLMTLSNVCSLN 60
Qy 61 TPARKKTGSHLILLOKICPRLYGPNTQACCSAKQVLSLEASITKALITRCPCASDN 120
Db 61 TPARKKTGSHLILLOKICPRLYGPNTQACCSAKQVLSLEASITKALITRCPCASDN 120
Qy 121 VNLHCNTPSNQSLFINTVVAQLGAGOLPAVVAEAPYQHSFAQSYDSCSRVRPAA 180
Db 121 VNLHCNTPSNQSLFINTVVAQLGAGOLPAVVAEAPYQHSFAQSYDSCSRVRPAA 180
Qy 181 ATLAAGTCMGVGSALCNARWLNFGQDTGNGIAPLDITFHLLEPCQAVSGIOPLINEGV 240
Db 181 ATLAAGTCMGVGSALCNARWLNFGQDTGNGIAPLDITFHLLEPCQAVSGIOPLINEGV 240
Qy 241 ARCNEQGDVATCSQODCAASCPALARPQALDSTYLGQMPGSLVLIILCSVPAVNTI 300
Db 241 ARCNEQGDVATCSQODCAASCPALARPQALDSTYLGQMPGSLVLIILCSVPAVNTI 300
Qy 301 ILVGFVAPARDRSKVNDPKGTSLSDKLSFTHTLLGQFQSGMTVVASWPTIIVLSV 360
Db 301 ILVGFVAPARDRSKVNDPKGTSLSDKLSFTHTLLGQFQSGMTVVASWPTIIVLSV 360
Qy 361 IPVALAAGLVFTELTTDPVELMSAPNSQARSEKAFHDOFGFPFRTOVLTAPNRSY 420
Db 361 IPVALAAGLVFTELTTDPVELMSAPNSQARSEKAFHDOFGFPFRTOVLTAPNRSY 420
Qy 421 RYDSSLGPRNFGIIDLLELLELQERLRLQVSPQAQNILODICTAPLPNDT 480
Db 421 RYDSSLGPRNFGIIDLLELLELQERLRLQVSPQAQNILODICTAPLPNDT 480
Qy 481 SLVDCCNISLLOFQNNRLLLTANOTLMGQSOVMDKHPLYCANAPLTFDQGTAL 540
Db 481 SLVDCCNISLLOFQNNRLLLTANOTLMGQSOVMDKHPLYCANAPLTFDQGTAL 540
Qy 541 SCNADYGAPVFPPLAIGYKGDYSEAEALIMFSLNNYPADGPRLOAKLMEAEFL 600
Db 541 SCNADYGAPVFPPLAIGYKGDYSEAEALIMFSLNNYPADGPRLOAKLMEAEFL 600
Qy 601 RAFORMAGNFQVTFARSLEDEINRTAEDLPFATSYIVFLYISLALGSSMSRV 660
Db 601 RAFORMAGNFQVTFARSLEDEINRTAEDLPFATSYIVFLYISLALGSSMSRV 660
Qy 661 MVDKATLIGGVAVVLGAVMAAMGFSSYLGRSSLVIIQVVPFLVLSVGADNIFIFVLE 720
Db 661 MVDKATLIGGVAVVLGAVMAAMGFSSYLGRSSLVIIQVVPFLVLSVGADNIFIFVLE 720
Qy 721 YQRLPRRPGPREVHIGRALGRVAPSMULCSLSEALICFPLGALTTPMPAVTTPALTSLAV 780
Db 721 YQRLPRRPGPREVHIGRALGRVAPSMULCSLSEALICFPLGALTTPMPAVTTPALTSLAV 780
Qy 781 IIDFLQMSAFVALLSLDSKROBASRLDYCCCVKPOELPPPGGEGILLGFQKAVAPFL 840
Db 781 IIDFLQMSAFVALLSLDSKROBASRLDYCCCVKPOELPPPGGEGILLGFQKAVAPFL 840
Qy 841 LHMITGVVLLFLALFGVSLYSMCHISVGLDOELAPKDSYLLDYFLPLNRYFEVGA 900
Db 841 LHMITGVVLLFLALFGVSLYSMCHISVGLDOELAPKDSYLLDYFLPLNRYFEVGA 900
Qy 901 YFVTTTGYNFSSEAGMNAICSSAGCNNSFTQKIQVATEPPEQSYLAIPASSWVDFIDW 960
Db 901 YFVTTTGYNFSSEAGMNAICSSAGCNNSFTQKIQVATEPPEQSYLAIPASSWVDFIDW 960
Qy 961 LTPSSCCRLYISGPNKDKFCPTVNSLNCCKMSITMGSVRPSVEQFHKYLPWFLNDRP 1020
Db 961 LTPSSCCRLYISGPNKDKFCPTVNSLNCCKMSITMGSVRPSVEQFHKYLPWFLNDRP 1020
Qy 1021 NIKCPKGLAAYSTSVNLTSDGOVLASRFMAVYHKPLKNSODYTEALPAARELANITADL 1080

```

```
Db      1021 NIKCPKGGIAAYSTVNLTSDDQVLA SRPMAVHKPLKNSQDYTEARARELANITADL 1080
Qy      1081 RKVPGTDPAFEVFPYTTINVFYEQYLITLPEGLFMLSCLVPTFAVSCLLGLDLSGGL 1140
Db      1081 RKVPGTDPAFEVFPYTTINVFYEQYLITLPEGLFMLSCLVPTFAVSCLLGLDLSGGL 1140
Qy      1141 NLLSTVMIIVDTVGVFMA LMDISYNAVSLINLVSAVMSVEFVSHITRSPALSTKPTWLER 1200
Db      1141 NLLSTVMIIVDTVGVFMA LMDISYNAVSLINLVSAVMSVEFVSHITRSPALSTKPTWLER 1200
Qy      1201 AKEATISMGSAVFAVAMNTNLPGLIIVLGLAKAQLIQIFFRNLNLTITLLGLHGVFLPV 1260
Db      1201 AKEATISMGSAVFAVAMNTNLPGLIIVLGLAKAQLIQIFFRNLNLTITLLGLHGVFLPV 1260
Qy      1261 ILSYVGPDPVPALALAEOKRAEBAVAVAWVASCPNHPRSVSTADNIYVNHSPFGS1KGAGA 1320
Db      1261 ILSYVGPDPVPALALAEOKRAEBAVAVAWVASCPNHPRSVSTADNIYVNHSPFGS1KGAGA 1320
Qy      1321 ISNFLPNNGROF 1332
Db      1321 ISNFLPNNGROF 1332
```

```
RESULT 4
US-10-736-769-4
/ Sequence 4, Application US/10736769
/ Publication No. US20040161838A1
/ GENERAL INFORMATION:
/ APPLICANT: Altman, Scott W
/ APPLICANT: Wang, Luquan
/ APPLICANT: Graziano, Michael
/ APPLICANT: Murgolo, Nick
/ TITLE OF INVENTION: NPC11 (NPC3) AND METHODS OF USE THEREOF
/ FILE REFERENCE: J01603-K3-US
/ CURRENT APPLICATION NUMBER: US/10/736,769
/ CURRENT FILING DATE: 2003-12-16
/ PRIOR APPLICATION NUMBER: 60/397,442
/ PRIOR FILING DATE: 2002-07-19
/ PRIOR APPLICATION NUMBER: 10/621,758
/ PRIOR FILING DATE: 2003-07-17
/ PRIOR APPLICATION NUMBER: 10/646,301
/ PRIOR FILING DATE: 2003-08-22
/ PRIOR APPLICATION NUMBER: 10/663,208
/ PRIOR FILING DATE: 2003-09-16
/ NUMBER OF SEQ ID NOS: 51
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 4
/ LENGTH: 1332
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-736-769-4
```

```
Query Match      100.0%; Score 6909; DB 4; Length 1332;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1332; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy      1 MABAGRGWMLWALLRLA QSEPYTTIHQPGYCAFYDECGKPELSGSLMTLSNVSCSN 60
Db      1 MABAGRGWMLWALLRLA QSEPYTTIHQPGYCAFYDECGKPELSGSLMTLSNVSCSN 60
Qy      61 TPARKITGDHLITLQKICPRLVYGPNTQA CCAKQVLSLEASISTTKALLTRCPACSDNF 120
Db      61 TPARKITGDHLITLQKICPRLVYGPNTQA CCAKQVLSLEASISTTKALLTRCPACSDNF 120
Qy      121 VNLHCNNTSPNOSLFINTTRVAQLGAGQLPAVVAEAFYQHSFAQSYDSCSRVVPAA 180
Db      121 VNLHCNNTSPNOSLFINTTRVAQLGAGQLPAVVAEAFYQHSFAQSYDSCSRVVPAA 180
Qy      181 ATLAVGTMGVYSALCA NQRMWLNFOGDTGNGLA PDIITFHLLPEQAVGSGIQPLNEGV 240
Db      181 ATLAVGTMGVYSALCA NQRMWLNFOGDTGNGLA PDIITFHLLPEQAVGSGIQPLNEGV 240
```

```
Qy      241 AHCNBSQGDVATCGCCODCA SCPAIARPOLDSTFYIGQMPSGLVITLLCSVAVVTI 300
Db      241 AHCNBSQGDVATCGCCODCA SCPAIARPOLDSTFYIGQMPSGLVITLLCSVAVVTI 300
Qy      301 LTVGFRVAPARDKSKMVPKCKGTSJSDKLSFSTHTLLAQFQCGMGTWVAWPLTILVSV 360
Db      301 LTVGFRVAPARDKSKMVPKCKGTSJSDKLSFSTHTLLAQFQCGMGTWVAWPLTILVSV 360
Qy      361 IPVVALAAGLVETELTTPDVELMSAPNSQARSEKA FHDQHFQGFPRTNQVILTAPNRSSY 420
Db      361 IPVVALAAGLVETELTTPDVELMSAPNSQARSEKA FHDQHFQGFPRTNQVILTAPNRSSY 420
Qy      421 RYDSILLGPKNSSGILDDILLELELOERLPHLQVSPBAQRNLSLODICAPLNPNPT 480
Db      421 RYDSILLGPKNSSGILDDILLELELOERLPHLQVSPBAQRNLSLODICAPLNPNPT 480
Qy      481 SLYDCCINSLOQFQNNRTLLLTANQTLMGQTSQVMDKHFLYCANNAPLTTRKDTALAL 540
Db      481 SLYDCCINSLOQFQNNRTLLLTANQTLMGQTSQVMDKHFLYCANNAPLTTRKDTALAL 540
Qy      541 SCMDYGA VPFPFLAIGGYKGDYSEAEALIMTFSLNYPAGDPRLAQAKLMEAFLEBM 600
Db      541 SCMDYGA VPFPFLAIGGYKGDYSEAEALIMTFSLNYPAGDPRLAQAKLMEAFLEBM 600
Qy      601 RAFORRMAGMFOYTFTAERSLEDEINRTTAEDLPFATSYIYIFLYISIALGSYSMSRV 660
Db      601 RAFORRMAGMFOYTFTAERSLEDEINRTTAEDLPFATSYIYIFLYISIALGSYSMSRV 660
Qy      661 MVDSKATLGLGVA VVLA VMAAMGFSSYLGRSSLVILQVVPFLVLSGADNIFLYLE 720
Db      661 MVDSKATLGLGVA VVLA VMAAMGFSSYLGRSSLVILQVVPFLVLSGADNIFLYLE 720
Qy      721 YQRLRRRPGEPREVAHIGRALGRVAPSMLLSISEAICFFLGALTMPAVRTFALTSGLA V 780
Db      721 YQRLRRRPGEPREVAHIGRALGRVAPSMLLSISEAICFFLGALTMPAVRTFALTSGLA V 780
Qy      781 IIDFLLQMSAFVALLS DSKQOEARLDVCCCVKQOELPPRQCGGLLLGFQKAYAPPL 840
Db      781 IIDFLLQMSAFVALLS DSKQOEARLDVCCCVKQOELPPRQCGGLLLGFQKAYAPPL 840
Qy      841 LHMITRGVVLTLFLALFGVSLYSMSCHISVGLDQELALPKDSTLIDYFLFLNRYFEVAGAV 900
Db      841 LHMITRGVVLTLFLALFGVSLYSMSCHISVGLDQELALPKDSTLIDYFLFLNRYFEVAGAV 900
Qy      901 YFVTTLGVNFSSEAGMNAICSSAGCNNSFTQKIQYATEFPBQSYLAI PASSWVDFTDW 960
Db      901 YFVTTLGVNFSSEAGMNAICSSAGCNNSFTQKIQYATEFPBQSYLAI PASSWVDFTDW 960
Qy      961 LTPSSCCRLYISGPKKDKFCPTVNSLNC LKNCMSITNGSVRPSVEQFHKYLPWFLNDRP 1020
Db      961 LTPSSCCRLYISGPKKDKFCPTVNSLNC LKNCMSITNGSVRPSVEQFHKYLPWFLNDRP 1020
Qy      1021 NIKCPKGGIAAYSTVNLTSDDQVLA SRPMAVHKPLKNSQDYTEARARELANITADL 1080
Db      1021 NIKCPKGGIAAYSTVNLTSDDQVLA SRPMAVHKPLKNSQDYTEARARELANITADL 1080
Qy      1081 RKVPGTDPAFEVFPYTTINVFYEQYLITLPEGLFMLSCLVPTFAVSCLLGLDLSGGL 1140
Db      1081 RKVPGTDPAFEVFPYTTINVFYEQYLITLPEGLFMLSCLVPTFAVSCLLGLDLSGGL 1140
Qy      1141 NLLSTVMIIVDTVGVFMA LMDISYNAVSLINLVSAVMSVEFVSHITRSPALSTKPTWLER 1200
Db      1141 NLLSTVMIIVDTVGVFMA LMDISYNAVSLINLVSAVMSVEFVSHITRSPALSTKPTWLER 1200
Qy      1201 AKEATISMGSAVFAVAMNTNLPGLIIVLGLAKAQLIQIFFRNLNLTITLLGLHGVFLPV 1260
Db      1201 AKEATISMGSAVFAVAMNTNLPGLIIVLGLAKAQLIQIFFRNLNLTITLLGLHGVFLPV 1260
Qy      1261 ILSYVGPDPVPALALAEOKRAEBAVAVAWVASCPNHPRSVSTADNIYVNHSPFGS1KGAGA 1320
Db      1261 ILSYVGPDPVPALALAEOKRAEBAVAVAWVASCPNHPRSVSTADNIYVNHSPFGS1KGAGA 1320
Qy      1321 ISNFLPNNGROF 1332
```

```
Db      1321  ISNPLPNNGRPF 1332
RESULT 5
US-10-239-316-9
; Sequence 9, Application US/10239316
; Publication No. US20030125253A1
; GENERAL INFORMATION:
; APPLICANT: TANIMAYA, Yoshio
; APPLICANT: KITA, Shunbun
; APPLICANT: SATOMI, Tomoko Komiyaama
; TITLE OF INVENTION: No. US20030125253a1e1 Protein, Process for Producing The Same And
; FILE REFERENCE: 2703USOP
; CURRENT APPLICATION NUMBER: US/10/239, 316
; PRIOR FILING DATE: 2002-09-19
; PRIOR APPLICATION NUMBER: PCT/JP01/02279
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: JP2000-088595
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 59
; SEQ ID NO 9
; LENGTH: 1332
; TYPE: PRT
; ORGANISM: Human
US-10-239-316-9

Query Match      99.8%; Score 6896; DB 4; Length 1332;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1330; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1  MAAAGLRGMLLMALLRLAASEPYTTIHQPGYCAFYDECGNPELSGLMTLSNVSCLSN 60
DB      1  MAAERGLRGMMLMLALLLAOSEPYTTIHQPGYCAFYDECGNPELSGLMTLSNVSCLSN 60
QY      61  TPARKITGDHLLILOKICPRLYTGPNTQACCSAQOLVLEASLSTTALTRCPACSDNF 120
DB      61  TPARKITGDHLLILOKICPRLYTGPNTQACCSAQOLVLEASLSTTALTRCPACSDNF 120
QY      121  VNHLCHNTCSPNOSLFINTRVNQLGAGQLPAVVAAYEAFYOHSAEBSYDSCSVRPAA 180
DB      121  VNHLCHNTCSPNOSLFINTRVNQLGAGQLPAVVAAYEAFYOHSAEBSYDSCSVRPAA 180
QY      121  VNHLCHNTCSPNOSLFINTRVNQLGAGQLPAVVAAYEAFYOHSAEBSYDSCSVRPAA 180
DB      121  VNHLCHNTCSPNOSLFINTRVNQLGAGQLPAVVAAYEAFYOHSAEBSYDSCSVRPAA 180
QY      181  ATLAAGTMCVYGSALCNAQRWLNFOGDTGNGLAFLDITFHLEBPGAVSGIQPLNEGV 240
DB      181  ATLAAGTMCVYGSALCNAQRWLNFOGDTGNGLAFLDITFHLEBPGAVSGIQPLNEGV 240
QY      181  ATLAAGTMCVYGSALCNAQRWLNFOGDTGNGLAFLDITFHLEBPGAVSGIQPLNEGV 240
DB      181  ATLAAGTMCVYGSALCNAQRWLNFOGDTGNGLAFLDITFHLEBPGAVSGIQPLNEGV 240
QY      241  ARCNEISQDDVATCSQDCAASCPAIPAPALDSTFYLGQMPGSLVLIILCSYFAVVTI 300
DB      241  ARCNEISQDDVATCSQDCAASCPAIPAPALDSTFYLGQMPGSLVLIILCSYFAVVTI 300
QY      301  LLYGFRVAPARDKSKMVDPKKGTSLSDKLSFTTTLTGQFQGGTVAASWPLTILVLSV 360
DB      301  LLYGFRVAPARDKSKMVDPKKGTSLSDKLSFTTTLTGQFQGGTVAASWPLTILVLSV 360
QY      301  LLYGFRVAPARDKSKMVDPKKGTSLSDKLSFTTTLTGQFQGGTVAASWPLTILVLSV 360
DB      301  LLYGFRVAPARDKSKMVDPKKGTSLSDKLSFTTTLTGQFQGGTVAASWPLTILVLSV 360
QY      361  IPVVALAAGLVFTLTTPVYELMSAPNSQARSEKAFHDQHPFRFNNOVILTPNRSY 420
DB      361  IPVVALAAGLVFTLTTPVYELMSAPNSQARSEKAFHDQHPFRFNNOVILTPNRSY 420
QY      421  RYDSLILGPKNFSGILDLLLELLELQERLRLHQVSPSAORNIISLQDICYAPLNDPT 480
DB      421  RYDSLILGPKNFSGILDLLLELLELQERLRLHQVSPSAORNIISLQDICYAPLNDPT 480
QY      421  RYDSLILGPKNFSGILDLLLELLELQERLRLHQVSPSAORNIISLQDICYAPLNDPT 480
DB      421  RYDSLILGPKNFSGILDLLLELLELQERLRLHQVSPSAORNIISLQDICYAPLNDPT 480
QY      481  SLVDCCINSLLQYQNNRTLLLTANQTLMGQTSQVDMKHFYLCANAPLTFKGTALAL 540
DB      481  SLVDCCINSLLQYQNNRTLLLTANQTLMGQTSQVDMKHFYLCANAPLTFKGTALAL 540
QY      541  SCMDYAPVPPFLAIGYKGYSEAEALMTSLSNNTYPAGDPRLLQAKLMEAPLEEM 600
DB      541  SCMDYAPVPPFLAIGYKGYSEAEALMTSLSNNTYPAGDPRLLQAKLMEAPLEEM 600
QY      601  RAFORMAQMFQVTFIERSLEDEINRTTADLPITATSYIVIFLYISLAGSYSSMSRV 660
DB      601  RAFORMAQMFQVTFIERSLEDEINRTTADLPITATSYIVIFLYISLAGSYSSMSRV 660
```

```
Db      601  RAFORMAQMFQVTFIERSLEDEINRTTADLPITATSYIVIFLYISLAGSYSSMSRV 660
QY      661  MVDSKATLGLGVAVVLGAVMAAMGFPSTYIGIRSSLVILQVPELVLSVGDNIPIFVLE 720
DB      661  MVDSKATLGLGVAVVLGAVMAAMGFPSTYIGIRSSLVILQVPELVLSVGDNIPIFVLE 720
QY      721  YORLPRRRGGPREVHIGRALGRVAPSMULCSLSSEALTFEFGALTPMAVNTFTALTSGLAV 780
DB      721  YORLPRRRGGPREVHIGRALGRVAPSMULCSLSSEALCFEFGALTPMAVNTFTALTSGLAV 780
QY      781  ILDFLLQMSAFVALLSLDSKROEASRLDVCCCVKPOELPPGQGEGLLGFQQAIVAPFL 840
DB      781  ILDFLLQMSAFVALLSLDSKROEASRLDVCCCVKPOELPPGQGEGLLGFQQAIVAPFL 840
QY      841  LHMITRGVLLFLALFGVSLYSWCHISVGLDELAPKDSYLLDYELFLNRYEAGAPV 900
DB      841  LHMITRGVLLFLALFGVSLYSWCHISVGLDELAPKDSYLLDYELFLNRYEAGAPV 900
QY      901  YFVTTLAGYNFSSRAGMAITCSAGCNNSFTQKIQVATERPEQSYLAIPASSWVDPTIDW 960
DB      901  YFVTTLAGYNFSSRAGMAITCSAGCNNSFTQKIQVATERPEQSYLAIPASSWVDPTIDW 960
QY      961  LTPSSCCRLYISGPNKDKPCPSTVNSLNCXKCMSTMGSVRPSVEQFHKYLPWFLNDRP 1020
DB      961  LTPSSCCRLYISGPNKDKPCPSTVNSLNCXKCMSTMGSVRPSVEQFHKYLPWFLNDRP 1020
QY      1021  NIKCPKGLAAVSTSVNLTSQGVLASRFMAHYHFKPLNSODYTEALRAARELANITADL 1080
DB      1021  NIKCPKGLAAVSTSVNLTSQGVLASRFMAHYHFKPLNSODYTEALRAARELANITADL 1080
QY      1081  RKVGTDPAREVPEPYTTITNVFYEQYLTILPEGLFMLSCLVPTFPAVSCLLGLDLSGGL 1140
DB      1081  RKVGTDPAREVPEPYTTITNVFYEQYLTILPEGLFMLSCLVPTFPAVSCLLGLDLSGGL 1140
QY      1141  NLSIWMILVDYGFALMDISYNAVSLINLVASVGSVRFVSHITSFALSTKPTWLER 1200
DB      1141  NLSIWMILVDYGFALMDISYNAVSLINLVASVGSVRFVSHITSFALSTKPTWLER 1200
QY      1201  AKEATISMSGAVFAGVAMTNPGLVYLGLAKAQLIQIFFRLNLLITLGLHLGLVPLV 1260
DB      1201  AKEATISMSGAVFAGVAMTNPGLVYLGLAKAQLIQIFFRLNLLITLGLHLGLVPLV 1260
QY      1261  ILSTYGPDVNPALALEQKRAEBAVAVVASCPNHPERSVSTADNIYNNHSFEGSIKAGA 1320
DB      1261  ILSTYGPDVNPALALEQKRAEBAVAVVASCPNHPERSVSTADNIYNNHSFEGSIKAGA 1320
QY      1321  ISNPLPNNGRPF 1332
DB      1321  ISNPLPNNGRPF 1332

RESULT 6
US-10-621-758A-44
; Sequence 44, Application US/10621758A
; Publication No. US20040093629A1
; GENERAL INFORMATION:
; APPLICANT: Altman, Scott W
; APPLICANT: Wang, Luquan
; APPLICANT: Graziano, Michael
; TITLE OF INVENTION: NPC1L1 (NPC3) AND METHODS OF USE THEREOF
; FILE REFERENCE: JB01603-K-US
; CURRENT APPLICATION NUMBER: US/10/621,758A
; PRIOR FILING DATE: 2003-07-17
; PRIOR APPLICATION NUMBER: 60/397,442
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 44
; LENGTH: 1359
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-621-758A-44
```

Query Match 99.5%; Score 6872.5; DB 4; Length 1359;
 Best Local Similarity 97.9%; Pred. No. 0;
 Matches 1330; Conservative 0; Mismatches 2; Indels 27; Gaps 1;

1 MAEAGRGWLWALLRLRLAQSPPYTTIHQGYCAFYDECGKNPELSGSLMTLSNVCLSN 60
 1 MAEAGRGWLWALLRLRLAQSPPYTTIHQGYCAFYDECGKNPELSGSLMTLSNVCLSN 60
 61 TPARKITGDHLLLOKICPRLYTGPNTQACSAKQVLSLEASLITKALLTRCPACSDNF 120
 61 TPARKITGDHLLLOKICPRLYTGPNTQACSAKQVLSLEASLITKALLTRCPACSDNF 120
 121 VNLHCNNTCSPNQSLFINVTRVAQLGAGQLPAVVAVEAFYQHSFAQSVDSCSRVVPAA 180
 121 VNLHCNNTCSPNQSLFINVTRVAQLGAGQLPAVVAVEAFYQHSFAQSVDSCSRVVPAA 180
 181 ATLAAGTCMGVYSALCNAQRMWLNFGQDGTGNGLAPLDITFHLLPEQAVGSGIQPLNEGV 240
 181 ATLAAGTCMGVYSALCNAQRMWLNFGQDGTGNGLAPLDITFHLLPEQAVGSGIQPLNEGV 240
 241 ARCNESQGDVATCSQDCASCPAARPAQALDSTFYLGQMPGSLVLIILCSVFAVVTI 300
 241 ARCNESQGDVATCSQDCASCPAARPAQALDSTFYLGQMPGSLVLIILCSVFAVVTI 300
 301 LTVGFVAPARDKSKNVDPKKGTSLSDKLSFSTHTTLTGQFPQMGWVAVSWPLTIIIVLSV 360
 301 LTVGFVAPARDKSKNVDPKKGTSLSDKLSFSTHTTLTGQFPQMGWVAVSWPLTIIIVLSV 360
 361 IPVVALAAGLVFPELTTDPELMSAPNSQARSEKAFHDOHFGFPFTNOVITLAPRRSSY 420
 361 IPVVALAAGLVFPELTTDPELMSAPNSQARSEKAFHDOHFGFPFTNOVITLAPRRSSY 420
 421 RYDSLILGKPNFSGILDLLELLELLELLELLELLELLELLELLELLELLELLELLELLEL 480
 421 RYDSLILGKPNFSGILDLLELLELLELLELLELLELLELLELLELLELLELLELLELLEL 480
 481 SLYDCINSLLOYFQNNRLLILLITANQTLMGQTSQVDMKHFLYCANAPLTFQDGTALAL 540
 481 SLYDCINSLLOYFQNNRLLILLITANQTLMGQTSQVDMKHFLYCANAPLTFQDGTALAL 540
 541 SCMAVDGAVFPPLAIGYKKGKDYSEAEALIMFSLNNYPAGBPRLAQAALMEAELEEM 600
 541 SCMAVDGAVFPPLAIGYKKGKDYSEAEALIMFSLNNYPAGBPRLAQAALMEAELEEM 600
 601 RAFORRMAGMFOYTFPAERSLDEINRTTAEDLPITATSYIVFLYISLALGSYSMSRY 660
 601 RAFORRMAGMFOYTFPAERSLDEINRTTAEDLPITATSYIVFLYISLALGSYSMSRY 660
 661 MYDSKATLIGLGVAVVLGAVMAAMGFPSSYIGIRSSLVLIQVBPVLVSVGADNIFIVLE 720
 661 MYDSKATLIGLGVAVVLGAVMAAMGFPSSYIGIRSSLVLIQVBPVLVSVGADNIFIVLE 720
 721 YQRLPRRPGBPBRVHIGRALGRVAPSMILCSLSEALCFPLGALTTPMAPARTFALTSGLAV 780
 721 YQRLPRRPGBPBRVHIGRALGRVAPSMILCSLSEALCFPLGALTTPMAPARTFALTSGLAV 780
 781 IIDFLQMSAFVALSLDSKROEASRLDYCCCVKPELPPPGQEGILLGFQKAXAPFL 840
 781 IIDFLQMSAFVALSLDSKROEASRLDYCCCVKPELPPPGQEGILLGFQKAXAPFL 840
 841 LHMITRGVULLFLALFGVLSVMCHISVGLDOELALPKDSYLLDYFLFLNRYFEVGAVP 900
 841 LHMITRGVULLFLALFGVLSVMCHISVGLDOELALPKDSYLLDYFLFLNRYFEVGAVP 900
 901 YFVTTLIGNPSSPAGNNAICSSAGCNPFSFTQKIQTATEPPEQSYALIPASSVNDPFIW 960
 901 YFVTTLIGNPSSPAGNNAICSSAGCNPFSFTQKIQTATEPPEQSYALIPASSVNDPFIW 960
 961 LTPSSCRLYISGPNKDKCPSTVNSLNLCKNCSITMGSVRPSVEQFHXYLPMFLNDRP 1020
 961 LTPSSCRLYISGPNKDKCPSTVNSLNLCKNCSITMGSVRPSVEQFHXYLPMFLNDRP 1020

1021 NIKCPKGLAAYSTSVNLTSDDQVL-----ASRFMAYH 1053
 1021 NIKCPKGLAAYSTSVNLTSDDQVL-----ASRFMAYH 1080
 1054 KPLKNSQDYTBALRAARELANITADLRKVPGETDPAFEVFPYITNVFEQYLTILPEGL 1113
 1054 KPLKNSQDYTBALRAARELANITADLRKVPGETDPAFEVFPYITNVFEQYLTILPEGL 1140
 1114 FMLSICLVPTFAVSCILLGLDRSGLLNLISVMIILVDTVGFMALWDISYNAVSLINLV 1173
 1141 FMLSICLVPTFAVSCILLGLDRSGLLNLISVMIILVDTVGFMALWDISYNAVSLINLV 1200
 1174 AYGMSEYFVSHITRSPALITKPTWLEBAKEATISMGSAFPAQVAMTNLEGIIVLGAKQ 1233
 1201 AVGMSEYFVSHITRSPALITKPTWLEBAKEATISMGSAFPAQVAMTNLEGIIVLGAKQ 1260
 1234 LIQIPEFRNLITLLGLHGVFLPVILSYGPDVNPALALQKRAEBAVAAVMAVASC 1293
 1261 LIQIPEFRNLITLLGLHGVFLPVILSYGPDVNPALALQKRAEBAVAAVMAVASC 1320
 1294 NHPSRVSTADNITYNHSFEFSIKGAGAISNPLPNNGRQF 1332
 1321 NHPSRVSTADNITYNHSFEFSIKGAGAISNPLPNNGRQF 1359

RESULT 7
 US-10-663-208A-44
 ; Sequence 44, Application US/10663208A
 ; Publication No. US20040132058A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Altman, Scott W
 ; APPLICANT: Wang, Luquan
 ; APPLICANT: Graziano, Michael
 ; APPLICANT: Murgolo, Nick
 ; TITLE OF INVENTION: NPC1L1 (NPC3) AND METHODS OF USE THEREOF
 ; FILE REFERENCE: JB01603K2 US
 ; CURRENT APPLICATION NUMBER: US/10/663,208A
 ; CURRENT FILING DATE: 2003-09-16
 ; PRIOR APPLICATION NUMBER: 60/397,442
 ; PRIOR FILING DATE: 2002-07-19
 ; PRIOR APPLICATION NUMBER: 10/621,758
 ; PRIOR FILING DATE: 2003-07-17
 ; PRIOR APPLICATION NUMBER: 10/646,301
 ; PRIOR FILING DATE: 2003-08-22
 ; NUMBER OF SEQ ID NOS: 50
 ; SOFTWARE: Patentin version 3.1
 ; SEQ ID NO 44
 ; LENGTH: 1359
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-663-208A-44

Query Match 99.5%; Score 6872.5; DB 4; Length 1359;
 Best Local Similarity 97.9%; Pred. No. 0;
 Matches 1330; Conservative 0; Mismatches 2; Indels 27; Gaps 1;

1 MAEAGRGWLWALLRLRLAQSPPYTTIHQGYCAFYDECGKNPELSGSLMTLSNVCLSN 60
 1 MAEAGRGWLWALLRLRLAQSPPYTTIHQGYCAFYDECGKNPELSGSLMTLSNVCLSN 60
 61 TPARKITGDHLLLOKICPRLYTGPNTQACSAKQVLSLEASLITKALLTRCPACSDNF 120
 61 TPARKITGDHLLLOKICPRLYTGPNTQACSAKQVLSLEASLITKALLTRCPACSDNF 120
 121 VNLHCNNTCSPNQSLFINVTRVAQLGAGQLPAVVAVEAFYQHSFAQSVDSCSRVVPAA 180
 121 VNLHCNNTCSPNQSLFINVTRVAQLGAGQLPAVVAVEAFYQHSFAQSVDSCSRVVPAA 180
 1261 LIQIPEFRNLITLLGLHGVFLPVILSYGPDVNPALALQKRAEBAVAAVMAVASC 1293
 1261 LIQIPEFRNLITLLGLHGVFLPVILSYGPDVNPALALQKRAEBAVAAVMAVASC 1320
 1294 NHPSRVSTADNITYNHSFEFSIKGAGAISNPLPNNGRQF 1332
 1321 NHPSRVSTADNITYNHSFEFSIKGAGAISNPLPNNGRQF 1359

Db 241 ARCNESQGDVATCSQDDCAASCPIARPOLDSTFYLQMPGSLVLIILICSVFAVVTI 300
 QY 301 LTVGFVPAARDKSKMNDPKKGTSLSDKLSFSTHTLLGQFQMGWVWASMPILITLVSV 360
 Db 301 LTVGFVPAARDKSKMNDPKKGTSLSDKLSFSTHTLLGQFQMGWVWASMPILITLVSV 360
 QY 361 IPVVALAAGVFTTELTPDVEILMSAPNSOARSEKAFHDOHGFPRFTNOVILTAENRSY 420
 Db 361 IPVVALAAGVFTTELTPDVEILMSAPNSOARSEKAFHDOHGFPRFTNOVILTAENRSY 420
 QY 421 RYDSLLGPKNFSGLIDLDLLELELOERLHLQVMSPEAORNI SLODI CYAPLNDNT 480
 Db 421 RYDSLLGPKNFSGLIDLDLLELELOERLHLQVMSPEAORNI SLODI CYAPLNDNT 480
 QY 481 SLVDCINSLLQYFQNNRTLLLTANQTLMGQTSQVDMKDFLYCANAPLTFKDGITALAL 540
 Db 481 SLVDCINSLLQYFQNNRTLLLTANQTLMGQTSQVDMKDFLYCANAPLTFKDGITALAL 540
 QY 541 SCMAVYGAVPFPFLAIGYKGDYSEABALIMTFSLNYPADGPRLAQAKLMEBAFLBEM 600
 Db 541 SCMAVYGAVPFPFLAIGYKGDYSEABALIMTFSLNYPADGPRLAQAKLMEBAFLBEM 600
 QY 601 RAFORMAAGMFOVTFMERSLEDEINRTAEDLPFATSYIVIFLYISLALGYSMSRV 660
 Db 601 RAFORMAAGMFOVTFMERSLEDEINRTAEDLPFATSYIVIFLYISLALGYSMSRV 660
 QY 661 MVDKATLGLGAVVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVV 720
 Db 661 MVDKATLGLGAVVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVV 720
 QY 721 YQRLPRRGEPREVHIGRALGRVAPSMULCSLSEALCEFLALTPMPAVRTFALTSGLAV 780
 Db 721 YQRLPRRGEPREVHIGRALGRVAPSMULCSLSEALCEFLALTPMPAVRTFALTSGLAV 780
 QY 781 ILDFLQMSAFVALLISLDSKQOASRLDVCCKVQOELPRPGQSGLLGFFQXAYAFEL 840
 Db 781 ILDFLQMSAFVALLISLDSKQOASRLDVCCKVQOELPRPGQSGLLGFFQXAYAFEL 840
 QY 841 LHMVITRGVALLFLALFGVSLYSMCHISVGLDQELAPKDSYLDYFLMLRYPFVGA 900
 Db 841 LHMVITRGVALLFLALFGVSLYSMCHISVGLDQELAPKDSYLDYFLMLRYPFVGA 900
 QY 901 YFVTTLLGNFSSEAGMNAICSSAGCNNSFTOKIYATEFPPOSYALIPASSWVDFIDW 960
 Db 901 YFVTTLLGNFSSEAGMNAICSSAGCNNSFTOKIYATEFPPOSYALIPASSWVDFIDW 960
 QY 961 LTPSSCCRLYISGPNKDFCSTVNSLNCNKCMSITMGSVRPSVEQFHXYLPWFLNDRP 1020
 Db 961 LTPSSCCRLYISGPNKDFCSTVNSLNCNKCMSITMGSVRPSVEQFHXYLPWFLNDRP 1020
 QY 1021 NIKCPKGLAAYSTSVNLTSDQVLT-----ASRPMAYH 1053
 Db 1021 NIKCPKGLAAYSTSVNLTSDQVLT-----ASRPMAYH 1053
 QY 1054 KPLKNSQDYTEALRAARELIANITADLRKVPCTDPAFEPYTTINVPFOYLITLPEEL 1113
 Db 1054 KPLKNSQDYTEALRAARELIANITADLRKVPCTDPAFEPYTTINVPFOYLITLPEEL 1113
 QY 1114 FMLSLCTVPTFAVSCLLGLDLSGLNLSTVMILVDVGFMLMDISYNAVSLINLVS 1173
 Db 1114 FMLSLCTVPTFAVSCLLGLDLSGLNLSTVMILVDVGFMLMDISYNAVSLINLVS 1173
 QY 1174 ANGMSEVFSHTTRSPALSTKPTWLERAKEATISMSAVFAGVAMTNLPGILVGLAKAO 1233
 Db 1174 ANGMSEVFSHTTRSPALSTKPTWLERAKEATISMSAVFAGVAMTNLPGILVGLAKAO 1233
 QY 1234 LIOIEFFRLNLLITLGLHGLVFLPVILSYGPDVNPALALBQKABEAVAAVWASCP 1293
 Db 1234 LIOIEFFRLNLLITLGLHGLVFLPVILSYGPDVNPALALBQKABEAVAAVWASCP 1293
 QY 1261 LIOIEFFRLNLLITLGLHGLVFLPVILSYGPDVNPALALBQKABEAVAAVWASCP 1320
 Db 1261 LIOIEFFRLNLLITLGLHGLVFLPVILSYGPDVNPALALBQKABEAVAAVWASCP 1320
 QY 1294 NHPRSVSTADNIYVNSHFSFGSIKGAGAINSNFLPNNGROF 1332
 Db 1294 NHPRSVSTADNIYVNSHFSFGSIKGAGAINSNFLPNNGROF 1332

Db 1321 NHPRSVSTADNIYVNSHFSFGSIKGAGAINSNFLPNNGROF 1359
 RESULT 8
 US-10-646-301A-44
 ; Sequence 44, Application US/10646301A
 ; Publication NO. US20040137467A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Altman, Scott W
 ; APPLICANT: Wang, Luquan
 ; APPLICANT: Graziano, Michael
 ; APPLICANT: Margolo, Nick
 ; TITLE OF INVENTION: NPC1L1 (NPC3) AND METHODS OF USE THEREOF
 ; FILE REFERENCE: J01603-K1-US
 ; CURRENT APPLICATION NUMBER: US/10/646,301A
 ; CURRENT FILING DATE: 2003-08-22
 ; PRIOR APPLICATION NUMBER: 60/397,442
 ; PRIOR FILING DATE: 2002-07-19
 ; PRIOR APPLICATION NUMBER: 10/621,758
 ; PRIOR FILING DATE: 2003-07-17
 ; NUMBER OF SEQ ID NOS: 50
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 44
 ; LENGTH: 1359
 ; TYPE: PRF
 ; ORGANISM: Homo sapiens
 US-10-646-301A-44
 Query Match 99.5%; Score 6872.5; DB 4; Length 1359;
 Best Local Similarity 97.9%; Pred. No. 0;
 Matches 1330; Conservative 0; Mismatches 2; Indels 27; Gaps 1;
 QY 1 MAEGLRGLMALLLRLAOSERTTHHPCYCAFYDECGNPLSGSLMTLSNVSCLSN 60
 Db 1 MAEGLRGLMALLLRLAOSERTTHHPCYCAFYDECGNPLSGSLMTLSNVSCLSN 60
 QY 61 TPARKITGDHLLLOKICPRLYTGPNTQACCSAKQVLSLEASLITLALTRCPACSNF 120
 Db 61 TPARKITGDHLLLOKICPRLYTGPNTQACCSAKQVLSLEASLITLALTRCPACSNF 120
 QY 121 VNLHCNHTCSPNQSLFINTVTAQLGAGOLPAVVAVEAFYQHSFAEODYSCSRVVPAA 180
 Db 121 VNLHCNHTCSPNQSLFINTVTAQLGAGOLPAVVAVEAFYQHSFAEODYSCSRVVPAA 180
 QY 121 VNLHCNHTCSPNQSLFINTVTAQLGAGOLPAVVAVEAFYQHSFAEODYSCSRVVPAA 180
 Db 121 VNLHCNHTCSPNQSLFINTVTAQLGAGOLPAVVAVEAFYQHSFAEODYSCSRVVPAA 180
 QY 181 ATLAAGTGCYGGGALCNAQRMNLPQDGTGNGALPLDITTHLLEBQAVSGIOPINAGV 240
 Db 181 ATLAAGTGCYGGGALCNAQRMNLPQDGTGNGALPLDITTHLLEBQAVSGIOPINAGV 240
 QY 181 ATLAAGTGCYGGGALCNAQRMNLPQDGTGNGALPLDITTHLLEBQAVSGIOPINAGV 240
 Db 181 ATLAAGTGCYGGGALCNAQRMNLPQDGTGNGALPLDITTHLLEBQAVSGIOPINAGV 240
 QY 241 ARCNESQGDVATCSQDDCAASCPIARPOLDSTFYLQMPGSLVLIILICSVFAVVTI 300
 Db 241 ARCNESQGDVATCSQDDCAASCPIARPOLDSTFYLQMPGSLVLIILICSVFAVVTI 300
 QY 301 LTVGFVPAARDKSKMNDPKKGTSLSDKLSFSTHTLLGQFQMGWVWASMPILITLVSV 360
 Db 301 LTVGFVPAARDKSKMNDPKKGTSLSDKLSFSTHTLLGQFQMGWVWASMPILITLVSV 360
 QY 361 IPVVALAAGVFTTELTPDVEILMSAPNSOARSEKAFHDOHGFPRFTNOVILTAENRSY 420
 Db 361 IPVVALAAGVFTTELTPDVEILMSAPNSOARSEKAFHDOHGFPRFTNOVILTAENRSY 420
 QY 421 RYDSLLGPKNFSGLIDLDLLELELOERLHLQVMSPEAORNI SLODI CYAPLNDNT 480
 Db 421 RYDSLLGPKNFSGLIDLDLLELELOERLHLQVMSPEAORNI SLODI CYAPLNDNT 480
 QY 481 SLVDCINSLLQYFQNNRTLLLTANQTLMGQTSQVDMKDFLYCANAPLTFKDGITALAL 540
 Db 481 SLVDCINSLLQYFQNNRTLLLTANQTLMGQTSQVDMKDFLYCANAPLTFKDGITALAL 540
 QY 541 SCMAVYGAVPFPFLAIGYKGDYSEABALIMTFSLNYPADGPRLAQAKLMEBAFLBEM 600
 Db 541 SCMAVYGAVPFPFLAIGYKGDYSEABALIMTFSLNYPADGPRLAQAKLMEBAFLBEM 600
 QY 601 RAFORMAAGMFOVTFMERSLEDEINRTAEDLPFATSYIVIFLYISLALGYSMSRV 660
 Db 601 RAFORMAAGMFOVTFMERSLEDEINRTAEDLPFATSYIVIFLYISLALGYSMSRV 660

Db 601 RAFORRMAGMFQVTFPAERSLDEINRTTAEDLPITATSYIVFLYISLALSYSSMSRV 660
Qy 661 MVDKATITGLGVAVVLGAVMAAMGFPSYLGIRSSLVIIQVVPFLVSVGADNIFIFVLE 720
Db 661 MVDKATITGLGVAVVLGAVMAAMGFPSYLGIRSSLVIIQVVPFLVSVGADNIFIFVLE 720
Qy 721 YORLPARRPGEPRBVHIGRALGRVAPBSMLCSISEAICFPLGALTTPMPAVRTFALTSGLAV 780
Db 721 YORLPARRPGEPRBVHIGRALGRVAPBSMLCSISEAICFPLGALTTPMPAVRTFALTSGLAV 780
Qy 781 IIDPFLQMSAFVALSLDSKROBASRLDVCCCVKPOELPPPOGSEGLLGFQKAYAPFL 840
Db 781 IIDPFLQMSAFVALSLDSKROBASRLDVCCCVKPOELPPPOGSEGLLGFQKAYAPFL 840
Qy 841 LHMITRGVLLFLALFGVSLYSKCHISVGLDQELALPKDSYLLDYFLFLNRYFEVGA 900
Db 841 LHMITRGVLLFLALFGVSLYSKCHISVGLDQELALPKDSYLLDYFLFLNRYFEVGA 900
Qy 901 YEVTTLGYNFSSBAGNNAICSSAGCNSFTQKIQYATEPPEQSYLAIPASSVWDPFIDM 960
Db 901 YEVTTLGYNFSSBAGNNAICSSAGCNSFTQKIQYATEPPEQSYLAIPASSVWDPFIDM 960
Qy 961 LTPSSCCRLYISGPNKDFCPSTVNSLNCCKMCSITMGSVRPSVEQFHKTYLPMFLNDRP 1020
Db 961 LTPSSCCRLYISGPNKDFCPSTVNSLNCCKMCSITMGSVRPSVEQFHKTYLPMFLNDRP 1020
Qy 1021 NIKCPKGLAAVSTVNLTSDDQVL-----ASRFMAVH 1053
Db 1021 NIKCPKGLAAVSTVNLTSDDQVL-----ASRFMAVH 1053
Qy 1054 KPLKNSQDYTEALRAARELANITADLRKVPCTDPAEVPFYTTNVPFYQYTLTILEGI 1113
Db 1054 KPLKNSQDYTEALRAARELANITADLRKVPCTDPAEVPFYTTNVPFYQYTLTILEGI 1113
Qy 1114 PMLSLCLVETPAVSCLLGLDLSRGLNLTSTWIIIVDTGFMALDISYNAVSLNLSV 1173
Db 1141 PMLSLCLVETPAVSCLLGLDLSRGLNLTSTWIIIVDTGFMALDISYNAVSLNLSV 1173
Qy 1174 AVGMSVEFVSHITRFAISTKPTWLERAKEATISMSAVFAGVAMTNLPGLIVLGLAKQ 1233
Db 1201 AVGMSVEFVSHITRFAISTKPTWLERAKEATISMSAVFAGVAMTNLPGLIVLGLAKQ 1260
Qy 1234 LIQIIEFFRLNLTITLGLHGLVFLPVIIISYGPDPVPALALEQKAEBAVAAVMVASC 1293
Db 1261 LIQIIEFFRLNLTITLGLHGLVFLPVIIISYGPDPVPALALEQKAEBAVAAVMVASC 1320
Qy 1294 NHPRSVSTADNTVYHNSFEGSIKAGAIENFLPNNGROF 1332
Db 1321 NHPRSVSTADNTVYHNSFEGSIKAGAIENFLPNNGROF 1359
RESULT 9
US-10-736-769-44
; Sequence 44, Application US/10736769
; Publication No. US20040161838A1
; GENERAL INFORMATION:
; APPLICANT: Altman, Scott W
; APPLICANT: Wang, Luquan
; APPLICANT: Graziano, Michael
; APPLICANT: Murgolo, Nick
; TITLE OF INVENTION: NPC1L1 (NPC3) AND METHODS OF USE THEREOF
; FILE REFERENCE: J01603-K3-US
; CURRENT FILING DATE: 2003-12-16
; PRIOR FILING DATE: 2003-12-16
; PRIOR FILING DATE: 2002-07-19
; PRIOR FILING DATE: 2002-07-19
; PRIOR FILING DATE: 2002-07-19
; PRIOR FILING DATE: 2003-07-17
; PRIOR FILING DATE: 2003-07-17
; PRIOR FILING DATE: 2003-08-22
; PRIOR FILING DATE: 2003-08-22
; PRIOR FILING DATE: 2003-09-16

; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 44
; LENGTH: 1359
; TYPE: PR
; ORGANISM: Homo sapiens
US-10-736-769-44
Query Match 99.5%; Score 6872.5; DB 4; Length 1359;
Best Local Similarity 97.9%; Pred. No. 0;
Matches 1330; Conservative 0; Mismatches 2; Indels 27; Gaps 1;
Qy 1 MAEAGRGWLMALRLAOSRPYTHIQPGCAFYDEGKRPISGMTSNVCSUN 60
Db 1 MAEAGRGWLMALRLAOSRPYTHIQPGCAFYDEGKRPISGMTSNVCSUN 60
Qy 61 TPARKITGDHLLLOKICPRLTYGPNTOCCSAKQVLSLEASITKALLTRCPACSDNF 120
Db 61 TPARKITGDHLLLOKICPRLTYGPNTOCCSAKQVLSLEASITKALLTRCPACSDNF 120
Qy 121 VNLHCHNTCSPNQSLFINTVVAQLAGQLPAVVAVEAFYQHSFABQSYDSCSRVRPAA 180
Db 121 VNLHCHNTCSPNQSLFINTVVAQLAGQLPAVVAVEAFYQHSFABQSYDSCSRVRPAA 180
Qy 181 ATLAVGTMGVYGSALCNARWLNFOGDTGNGLAPLDITFHLLEBGOANGSIOPLNEGV 240
Db 181 ATLAVGTMGVYGSALCNARWLNFOGDTGNGLAPLDITFHLLEBGOANGSIOPLNEGV 240
Qy 241 ARCNSQGDVATCSCODCAACPAIARPOALDSTFYLGOMGSLVLIILCSVPAVVTI 300
Db 241 ARCNSQGDVATCSCODCAACPAIARPOALDSTFYLGOMGSLVLIILCSVPAVVTI 300
Qy 301 LLVGRVAPARDKSKVDPKGTSLSDKLSFSTHTLLGQFQCGKGTWASWPLTIVLSV 360
Db 301 LLVGRVAPARDKSKVDPKGTSLSDKLSFSTHTLLGQFQCGKGTWASWPLTIVLSV 360
Qy 361 IPVVALAGLVETELTTPVVELMSAPNSQABSEKAFHDOHPFPRTNOVLIITAPRSSY 420
Db 361 IPVVALAGLVETELTTPVVELMSAPNSQABSEKAFHDOHPFPRTNOVLIITAPRSSY 420
Qy 421 RYDSLLGPKNSGLIDDLLELELOERLHLQVSPSEARNTSLDIOICAPLPNDT 480
Db 421 RYDSLLGPKNSGLIDDLLELELOERLHLQVSPSEARNTSLDIOICAPLPNDT 480
Qy 481 SLYDCCINSLOYFQNNRTLLLTANQTMGOTSQVDWKDHLFYCANAPLTKDGTALAL 540
Db 481 SLYDCCINSLOYFQNNRTLLLTANQTMGOTSQVDWKDHLFYCANAPLTKDGTALAL 540
Qy 541 SCMAOVGAPVFPFLAIGYKGYSEAEALIMTFSLNNYPAGDPRLAQKMEBAFLEEM 600
Db 541 SCMAOVGAPVFPFLAIGYKGYSEAEALIMTFSLNNYPAGDPRLAQKMEBAFLEEM 600
Qy 601 RAFORRMAGMFQVTFPAERSLDEINRTTAEDLPITATSYIVFLYISLALSYSSMSRV 660
Db 601 RAFORRMAGMFQVTFPAERSLDEINRTTAEDLPITATSYIVFLYISLALSYSSMSRV 660
Qy 661 MVDKATITGLGVAVVLGAVMAAMGFPSYLGIRSSLVIIQVVPFLVSVGADNIFIFVLE 720
Db 661 MVDKATITGLGVAVVLGAVMAAMGFPSYLGIRSSLVIIQVVPFLVSVGADNIFIFVLE 720
Qy 721 YORLPARRPGEPRBVHIGRALGRVAPBSMLCSISEAICFPLGALTTPMPAVRTFALTSGLAV 780
Db 721 YORLPARRPGEPRBVHIGRALGRVAPBSMLCSISEAICFPLGALTTPMPAVRTFALTSGLAV 780
Qy 781 IIDPFLQMSAFVALSLDSKROBASRLDVCCCVKPOELPPPOGSEGLLGFQKAYAPFL 840
Db 781 IIDPFLQMSAFVALSLDSKROBASRLDVCCCVKPOELPPPOGSEGLLGFQKAYAPFL 840
Qy 841 LHMITRGVLLFLALFGVSLYSKCHISVGLDQELALPKDSYLLDYFLFLNRYFEVGA 900
Db 841 LHMITRGVLLFLALFGVSLYSKCHISVGLDQELALPKDSYLLDYFLFLNRYFEVGA 900
Qy 901 YEVTTLGYNFSSBAGNNAICSSAGCNSFTQKIQYATEPPEQSYLAIPASSVWDPFIDM 960


```

Db      901  YFTTIGYNSSSAGMNAICSSAGCNNSFTQKIQTATEPPEQSYLAIPASSWDDPDIW 960
Qy      961  LTPSSCCRLYISGPNKDKFCPTVNSLNCJKNCSITMGSVSPSVEQFHXYLPMFLNDRP 1020
Db      961  LTPSSCCRLYISGPNKDKFCPTVNSLNCJKNCSITMGSVSPSVEQFHXYLPMFLNDRP 1020
Qy      1021  NIKCPKGLAAYSTSVNLTSDGVL-----ASRFMAVH 1053
Db      1021  NIKCPKGLAAYSTSVNLTSDGVLDTVALISPRLEYSGTISAHCNLYLSDASRFMAVH 1080
Qy      1054  KPLKNSODTTEALRAARELANITADLRKPGTDPAREVEPPYTTINVFYQYLTILPEGL 1113
Db      1081  KPLKNSODTTEALRAARELANITADLRKPGTDPAREVEPPYTTINVFYQYLTILPEGL 1140
Qy      1114  FMSLCLVPTFAVSCLLGLDLSGLNLTLSTVILVDTGFMALMDISYNAVSLINLV 1173
Db      1141  FMSLCLVPTFAVSCLLGLDLSGLNLTLSTVILVDTGFMALMDISYNAVSLINLV 1200
Qy      1174  AVGMSVEPVSHITRSFAISTKPTWLEBAKEATISMSGSAVAGVAMTNLPGLIYLGLAKQ 1233
Db      1201  AVGMSVEPVSHITRSFAISTKPTWLEBAKEATISMSGSAVAGVAMTNLPGLIYLGLAKQ 1260
Qy      1234  LIOIFPFRNLTLTLGLHGLVFLPYILSYGPDVNPALAEQKRAEEVAAMVASC 1293
Db      1261  LIOIFPFRNLTLTLGLHGLVFLPYILSYGPDVNPALAEQKRAEEVAAMVASC 1320
Qy      1294  NHRSRVSTADNIYVNSFEFSIKGAGAINFLPNNGROF 1332
Db      1321  NHRSRVSTADNIYVNSFEFSIKGAGAINFLPNNGROF 1359

RESULT 10
US-10-450-763-53052
; Sequence 53052, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 53052
; LENGTH: 1344
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (901)..(917)
; OTHER INFORMATION: OESTROGEN RECEPTOR SIGNATURE domain identified by eMATRIX.
; OTHER INFORMATION: accession number PR00543D, p-value=1.355e-09, raw score of 10.87
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1344)
; OTHER INFORMATION: Xaa = X or * as defined in Table 2
US-10-450-763-53052

Query Match          94.6%; Score 6536; DB 5; Length 1344;
Best Local Similarity 96.7%; Pred. No. 0;
Matches 1271; Conservative 7; Mismatches 21; Indels 16; Gaps 4;

Qy      1  MEAGRGWMLTALRLIAGSEPYTTIHQGYCAFYDECGKNPELSSGLMTLSNVSLN 60
Db      1  MEAGRGWMLTALRLIAGSEPYTTIHQGYCAFYDECGKNPELSSGLMTLSNVSLN 60
Qy      61  TPAKRTGCHLILLOKICRPLLYTGPNTOACCSAKOLVSLPASISTKALLTRCPAGSDNF 120
Db      61  TPAKRTGCHLILLOKICRPLLYTGPNTOACCSAKOLVSLPASISTKALLTRCPAGSDNF 120
Qy      121  VNIHCHNTSPNOSLFINTRVAGOLPAVVAEAFQHSFAEOSYDSCGRVPA 180
Db      121  VNIHCHNTSPNOSLFINTRVAGOLPAVVAEAFQHSFAEOSYDSCGRVPA 180
Qy      181  ATLAATMGCGVGSALCNARWLNFOGDTGNGIAPLIDITPHLLPEQAVSGIOPLNEGV 240
Db      181  ATLAATMGCGVGSALCNARWLNFOGDTGNGIAPLIDITPHLLPEQAVSGIOPLNEGV 240
Qy      241  ARCNSQGDVATCSQDCAACP--ALARPALDSTFYLGQMPGSLVLIILICSVAVV 298
Db      241  ARCNSQGDVATCSQDCAACP--ALARPALDSTFYLGQMPGSLVLIILICSVAVV 298
Qy      299  TILVGFVAPADKSKMDPKKSTISDKLSSTHTLLGQFPGMGTVASMPLTLYL 358
Db      299  TILVGFVAPADKSKMDPKKSTISDKLSSTHTLLGQFPGMGTVASMPLTLYL 358
Qy      359  SVIPVVALAAGLVFTELTTDPVELMSAPNSQASEKAFDQHPGPFRTNQVILTAENRS 418
Db      359  SVIPVVALAAGLVFTELTTDPVELMSAPNSQASEKAFDQHPGPFRTNQVILTAENRS 418
Qy      419  SYRYDSSLGPKNFGSILDLILBELIOERLHHQVSPBAQRNISLODICYAPLND 478
Db      419  SYRYDSSLGPKNFGSILDLILBELIOERLHHQVSPBAQRNISLODICYAPLND 478
Qy      479  NTSLYOCCNSLIQYONNRRTLLLTANOTMGQTSQVMKHFLICANAPLTFKQGTAL 538
Db      479  NTSLYOCCNSLIQYONNRRTLLLTANOTMGQTSQVMKHFLICANAPLTFKQGTAL 538
Qy      539  ALSCMADYGAPEPFLAIGYKGDYSEABALIMTSLNNYPADPRLAQAKLMEAEFL 598
Db      539  ALSCMADYGAPEPFLAIGYKGDYSEABALIMTSLNNYPADPRLAQAKLMEAEFL 598
Qy      599  EKRAFORMAAGMFOVTFMABERSLEDEINRTTAEPLIPATSYVILFYILSLAGSYSSWS 658
Db      599  EKRAFORMAAGMFOVTFMABERSLEDEINRTTAEPLIPATSYVILFYILSLAGSYSSWS 658
Qy      659  RWVWDSKATLGLGAVAVVGAWMAAGFSYIGIRSVILOYVPLVLSVGADNIFIV 718
Db      659  RWVWDSKATLGLGAVAVVGAWMAAGFSYIGIRSVILOYVPLVLSVGADNIFIV 718
Qy      719  LEYQRLPRRPGEBREHIGRALGRVAPSMULCSLSAICFFLGALTPMRAVRFTALTSG 778
Db      719  LEYQRLPRRPGEBREHIGRALGRVAPSMULCSLSAICFFLGALTPMRAVRFTALTSG 778
Qy      779  AVILDFLOMSAFVALLSLDSKQASRLDVCCVCPQELPPPGQEGILLGFFQKAYAP 838
Db      779  AVILDFLOMSAFVALLSLDSKQASRLDVCCVCPQELPPPGQEGILLGFFQKAYAP 838
Qy      839  FLHMTTRGVV---LLFLALPGVELLYSMCHISVGLDDELALPKOSYLLDYFLFNRYF 894
Db      839  FLHMTTRGVV---LLFLALPGVELLYSMCHISVGLDDELALPKOSYLLDYFLFNRYF 894
Qy      895  EVGAPYFYTTLGYNFSSEAGMNAICSSAGCNNSFTQKIQTATEPPEQSYLAIPASSW 954
Db      895  EVGAPYFYTTLGYNFSSEAGMNAICSSAGCNNSFTQKIQTATEPPEQSYLAIPASSW 954
Qy      955  DPEIDWLTPSSCCRLYISGPNKDKFCPTVNSLNCJKNCSITMGSVSPSVEQFHXYLPW 1014
Db      955  DPEIDWLTPSSCCRLYISGPNKDKFCPTVNSLNCJKNCSITMGSVSPSVEQFHXYLPW 1014
Qy      1015  FLNDRPNIKCPKGLAAYSTSVNLTSDGVLASRFMAVHKPLKNSODTTEALRAAREL 1074
Db      1015  FLNDRPNIKCPKGLAAYSTSVNLTSDGVLASRFMAVHKPLKNSODTTEALRAAREL 1074
Qy      1075  NITADLRKVPGTDPAREVEPPYTTINVFYQYLTILPEGLFMSLCLVPTFAVSCLLGLD 1134
Db      1075  NITADLRKVPGTDPAREVEPPYTTINVFYQYLTILPEGLFMSLCLVPTFAVSCLLGLD 1134
Qy      1135  LRSGLNLTLSTVILVDTGFMALMDISYNAVSLINLVSAVGMSVEPVSHITRSFAISTK 1194
```

```

Db      1139  LRSGLNLSIWMIVDTYGFALWGI STNAVSLNLSAVGMSVEFVSHIRSFALSTK 1198
Qy      1195  PTWBRRAKATISMSGSAFVAGVAMTNLPGLVLVLGLAKAQLIOIPEFRMLNLTLLGLLHG 1254
Db      1199  PWTBRRAKATISMSGSAFVAGVAMTNLPGLVLVLGLAKAQLIOIPEFRMLNLTLLGLLHG 1258
Qy      1255  LVFLPVILSYGPDVNPALALBOKRAEZA-----VAAMVASCPNHRPVST 1301
Db      1259  LVFLPVILSYGPDVNPALALBOKRAEAGSGSHGSPALQITPPESPQLTTSMT 1313

RESULT 11
US-10-621-758A-2
; Sequence 2, Application US/10621758A
; Publication No. US20040093629A1
; GENERAL INFORMATION:
; APPLICANT: Altman, Scott W
; APPLICANT: Wang, Luquan
; APPLICANT: Graziano, Michael
; APPLICANT: Murgolo, Nick
; TITLE OF INVENTION: NPC1L1 (NPC3) AND METHODS OF USE THEREOF
; FILE REFERENCE: JB01603-K-US
; CURRENT APPLICATION NUMBER: US/10/621,758A
; CURRENT FILING DATE: 2003-07-17
; PRIOR APPLICATION NUMBER: 60/397,442
; PRIOR FILING DATE: 2002-07-19
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 1331
; TYPE: PRF
; ORGANISM: Rattus sp.
US-10-621-758A-2

Query Match      78.5%; Score 5421.5; DB 4; Length 1331;
Best Local Similarity 77.8%; Pred. No. 0;
Matches 1039; Conservative 120; Mismatches 169; Indels 7; Gaps 6;

Qy      1  MEAGLRGWLMAALLRLAOSBPYTTIHQGYCAFYDECGKPELSGSLMTLSNVASLSN 60
Db      1  MAAMU-GWLMNALLSAAOGELYTPKHAQVCTFEBCGKPELSGSLTSNVASLSN 59
Qy      61  TEARKITGDHLLQKICPRLYTGPTQ-ACCSAKOLVSLBASLITKALLTRCPACSDN 119
Db      60  TPAHNTYGHLLAQICPRLYNGPTTACCGTKQLSLBSMSITKALLTRCPACSDN 119
Qy      120  FVNLHNTNCSPPQSLFIVNTRVAQUGAQLPAVVA YEAFYQHSFAEGSYDSCRYRVA 179
Db      120  FVSLHCHNTCSPPQSLFIVNTRVERGABEPRAVVA YEAFYQHSFAEKSYESCSQVRI 179
Qy      180  AATLVATGTCGYGSAALCAQORWLNFGQGTNGLAFLDITPHLJEGQAVSGIOPLNG 239
Db      180  AASLAVGSMCGYGSALCAQORWLNFGQGTNGLAFLDITPHLJEGQAVSGIOPLNG 239
Qy      240  VARCNESQGDVATCSQDCAASCRAIARPAALDFTYLGOMGSLVLIILCSVAVYT 299
Db      240  IAPCNESQGDVATCSQDCAASCRAIARPAALDFTYLGOMGSLVLIILCSVAVYT 299
Qy      300  ILLVGRVAPARDKSMVDPKGTSLSDKLSFSTHTLLGQFQGMGTWASNPITLIVS 359
Db      300  AVLVRARVVSNNRKNKAEQPOEAPKLRPHKHLSPHTILRFQNMGTIRVAVSPLTYLAS 359
Qy      360  VLPVVALAGLVFTELTTPVEVLSAPNSQARSEKAFHNDHGFPRFTNOVLTAPNRS 419
Db      360  FIVVVALAGLVFTELTTPVEVLSAPNSQARSEKAFHNDHGFPRFTNOVLTAPNRS 419
Qy      420  YRYDSLILGPKNFSGILDLLELELEBERLHLQVMSPEAQRNLSLODICVAPLPNP 479
Db      420  YRYDSLILGPKNFSGILDLLELELEBERLHLQVMSPEAQRNLSLODICVAPLPNP 479
Qy      480  TSLYDCINSLLQYFQNNRTLLLTANQTLMGQTSQVDMKHFLYCANAPLTFKQGTALA 539
Db      480  TSLYDCINSLLQYFQNNRTLLLTANQTLMGQTSQVDMKHFLYCANAPLTFKQGTALA 539

```

```

Db      480  TSLYDCINSLLQYFQNNRTLLLTANQTLMGQTSQVDMKHFLYCANAPLTFKQGTALA 539
Qy      540  LSCMADYGAVPFPLAIGYKGDYSEAEALIMFESLNNYPAGDERLAQAKMEBAFEE 599
Db      540  LSCMADYGAVPFPLAIGYKGDYSEAEALIMFESLNNYPAGDERLAQAKMEBAFEE 599
Qy      600  MEAFORMAQMTFTFERSLEDEINRTAEDLPFATSYIVITLYSLALGYSMSR 659
Db      600  MEAFORMAQMTFTFERSLEDEINRTAEDLPFATSYIVITLYSLALGYSMSR 659
Qy      660  MESFORNTSDKQVAFSAERSLEDEINRTIODELVFVSVYIVELYSLALGYSRCS 719
Db      660  MESFORNTSDKQVAFSAERSLEDEINRTIODELVFVSVYIVELYSLALGYSRCS 719
Qy      720  EYORLPRRBPGEVHIGRALGRVAPSMULCSLSBAICFPGLALTMPAVRTFALTSG 779
Db      720  EYORLPRRBPGEVHIGRALGRVAPSMULCSLSBAICFPGLALTMPAVRTFALTSG 779
Qy      780  VILDFLQMSAFVALLSLSKQEASRLDVCCCVKQDELPPRGQEGILLGFQKAYAPF 839
Db      780  VILDFLQMSAFVALLSLSKQEASRLDVCCCVKQDELPPRGQEGILLGFQKAYAPF 839
Qy      840  LLMWITRGVLLFPALPGVSLYSWCHISVGLDQELAPKOSYLDYFLFLNRYEVA 899
Db      840  LLMWITRGVLLFPALPGVSLYSWCHISVGLDQELAPKOSYLDYFLFLNRYEVA 899
Qy      900  VYFVTTLLGYNFSSEAGMNAICSSAGCNPFSTOKIQVATEPEQSYLAI PASSW 959
Db      900  VYFVTTLLGYNFSSEAGMNAICSSAGCNPFSTOKIQVATEPEQSYLAI PASSW 959
Qy      960  WITP-SSCCRLYISGPNKDKFCPTSVNSLNCIKNMSITMGSVRSVQOFKXLYL 1018
Db      960  WITP-SSCCRLYISGPNKDKFCPTSVNSLNCIKNMSITMGSVRSVQOFKXLYL 1018
Qy      1019  RPNIKCPKGLAAAYSTVNLSDGOVLASRFMAHYKPKJNSODYTEALRAARELAN 1078
Db      1020  RPNIKCPKGLAAAYSTVNLSDGOVLASRFMAHYKPKJNSODYTEALRAARELAN 1078
Qy      1079  DLKAVPGTDPAFEPVPTITVNFYEOYLITLPEGLFMLSCLVPTFAVSCULLGL 1138
Db      1080  DLKAVPGTDPAFEPVPTITVNFYEOYLITLPEGLFMLSCLVPTFAVSCULLGL 1138
Qy      1139  LNLNLSIWMIVDTYGFALWGI STNAVSLNLSAVGMSVEFVSHIRSFALSTK 1198
Db      1140  LNLNLSIWMIVDTYGFALWGI STNAVSLNLSAVGMSVEFVSHIRSFALSTK 1198
Qy      1199  ERAKATISMSGSAFVAGVAMTNLPGLVLVLGLAKAQLIOIPEFRMLNLTLLGL 1258
Db      1200  ERAKATISMSGSAFVAGVAMTNLPGLVLVLGLAKAQLIOIPEFRMLNLTLLGL 1258
Qy      1259  PVILSYGPDVNPALALBOKRAEBAVVAAMVASCPNHRPVSTADNIVYNSHFE 1317
Db      1260  PVILSYGPDVNPALALBOKRAEBAVVAAMVASCPNHRPVSTADNIVYNSHFE 1317
Qy      1318  AGAISNPLPNNGROF 1332
Db      1317  ANAARSSLPKSDOKF 1331

RESULT 12
US-10-663-208A-2
; Sequence 2, Application US/10663208A
; Publication No. US20040132058A1
; GENERAL INFORMATION:
; APPLICANT: Altman, Scott W
; APPLICANT: Wang, Luquan
; APPLICANT: Graziano, Michael
; APPLICANT: Murgolo, Nick
; TITLE OF INVENTION: NPC1L1 (NPC3) AND METHODS OF USE THEREOF
; FILE REFERENCE: JB01603K2 US
; CURRENT APPLICATION NUMBER: US/10/663,208A
; CURRENT FILING DATE: 2003-09-16
; PRIOR APPLICATION NUMBER: 60/397,442

```

; PRIOR FILING DATE: 2002-07-19
 ; PRIOR APPLICATION NUMBER: 10/621,758
 ; PRIOR FILING DATE: 2003-07-17
 ; PRIOR APPLICATION NUMBER: 10/646,301
 ; PRIOR FILING DATE: 2003-08-22
 ; NUMBER OF SEQ ID NOS: 50
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 2
 ; LENGTH: 1331
 ; TYPE: PR1
 ; ORGANISM: Rattus sp.
 ; US-10-663-208A-2

Query Match 78.5%; Score 5421.5; DB 4; Length 1331;
 Best Local Similarity 77.8%; Pred. No. 0;
 Matches 1039; Conservative 120; Mismatches 169; Indels 7; Gaps 6;

QY 1 MAEAGRGWLLMALLRLAOSBPTTTHQPGYCAFYDECGKNPBLSGSLMTLSNVSCLSN 60
 DB 1 MAEAGRGWLLMALLRLAOSBPTTTHQPGYCAFYDECGKNPBLSGSLMTLSNVSCLSN 59
 QY 61 TPAARKITGDHLLILOKICPRLYNGPNTQ-ACCSAKOIVSLBASISITKALLTRCPACSDN 119
 DB 60 TPAARKITGDHLLILOKICPRLYNGPNTTACCSTKOLLSLESSWSITKALLTRCPACSDN 119
 QY 120 FVNLHCNTCSPNOSLFINTVTRVAQLAGOLPAVAVAYEAFYOHSPAEQSDSCSRVAPA 179
 DB 120 FVNLHCNTCSPNOSLFINTVTRVAGERAPVAVAYEAFYOHSPAEQSDSCSRVAPA 179
 QY 180 AATLAVGTGCVYGSALCAQRMWIFQDGTNGIAPLDITFHLLPEQAVSGIOPINBG 239
 DB 180 AATLAVGTGCVYGSALCAQRMWIFQDGTNGIAPLDITFHLLPEQAVSGIOPINBG 239
 QY 240 VACNESQGDVATCSCODCAAPARQALDSTFVLQMGPSVLIIILCSVEFVVT 299
 DB 240 VACNESQGDVATCSCODCAAPARQALDSTFVLQMGPSVLIIILCSVEFVVT 299
 QY 300 ILVGFVAPADKSKVNDPKGTSLSDKLSFTHTLLIGOFFQMGMTWVAMPLTIIIVLS 359
 DB 300 ILVGFVAPADKSKVNDPKGTSLSDKLSFTHTLLIGOFFQMGMTWVAMPLTIIIVLS 359
 QY 360 VIVVVALAAGLVFTLTTPDVEIWSAPNSQARSEKAFHDOHFGDFPFTNOVILTAPESS 419
 DB 360 VIVVVALAAGLVFTLTTPDVEIWSAPNSQARSEKAFHDOHFGDFPFTNOVILTAPESS 419
 QY 420 YKVDLSLLGKNSFGIISLDFLELLELQERLRLQVWSPAEERNISLODIQVAPLMPYN 479
 DB 420 YKVDLSLLGKNSFGIISLDFLELLELQERLRLQVWSPAEERNISLODIQVAPLMPYN 479
 QY 480 TSIYDCCINSILOFQONRTLLLTANOTLMGOTSQVDMKOHFLYCANAPLTFDGTALA 539
 DB 480 TSIYDCCINSILOFQONRTLLLTANOTLMGOTSQVDMKOHFLYCANAPLTFDGTALA 539
 QY 540 LSCADYAGVFPFLAIGYKGDYSEAEALIMTFSLNNYPAGDPLLAQAKLMEAELEB 599
 DB 540 LSCADYAGVFPFLAIGYKGDYSEAEALIMTFSLNNYPAGDPLLAQAKLMEAELEB 599
 QY 600 MRAFORMAGMFOVTFPAERSLEDEINKTTABDLPITATSTYIVFLYISLALGYSKSR 659
 DB 600 MRAFORMAGMFOVTFPAERSLEDEINKTTABDLPITATSTYIVFLYISLALGYSKSR 659
 QY 660 VMTDSKATLIGVAVVILGAVMAMGFPSTYIGIRSSIVIIQVAVFLVSVGADNIFIVL 719
 DB 660 VMTDSKATLIGVAVVILGAVMAMGFPSTYIGIRSSIVIIQVAVFLVSVGADNIFIVL 719
 QY 720 EYORLPERRPBPBPREVHIGRALGRVAPSMLLCSLESEACFFLGALTPMPAVTFTALSGLA 779
 DB 720 EYORLPERRPBPBPREVHIGRALGRVAPSMLLCSLESEACFFLGALTPMPAVTFTALSGLA 779
 QY 780 VIIDFLLQMSAFVALLSLDSKROASRLDVCCYKQDELPPRGQEGILLGFPQKAYAPF 839
 DB 780 VIIDFLLQMSAFVALLSLDSKROASRLDVCCYKQDELPPRGQEGILLGFPQKAYAPF 839

QY 840 LHMTRGVALLFLALFGVSLYSKCHISVGLDOELALPDOSYLLDYFLUNRYFEVGP 899
 DB 840 LHMTRGVALLFLALFGVSLYSKCHISVGLDOELALPDOSYLLDYFLUNRYFEVGP 899
 QY 900 VYFVTTLGVNPSBAGNNAICSSAGCNPFSTQIOYATPEPBGSYALTPASSVNDPFD 959
 DB 900 VYFVTTLGVNPSBAGNNAICSSAGCNPFSTQIOYATPEPBGSYALTPASSVNDPFD 959
 QY 960 WLP-SSCCRLYISGPNKDKFCSTVNSLNLCKNCSITWGSVAPSVYQFHKYLPWFLND 1018
 DB 960 WLP-SSCCRLYISGPNKDKFCSTVNSLNLCKNCSITWGSVAPSVYQFHKYLPWFLND 1018
 QY 1019 PENIKCPKGLAAYSTVNSLTSQOVLASFPMAYHFKLNSQDYTELARARELANITA 1078
 DB 1019 PENIKCPKGLAAYSTVNSLTSQOVLASFPMAYHFKLNSQDYTELARARELANITA 1078
 QY 1079 DLRVPGTDPAFEFPPPTINNVFEQVLTLPBGLFMLSCLVPTFVSCILLGLDURSG 1138
 DB 1079 DLRVPGTDPAFEFPPPTINNVFEQVLTLPBGLFMLSCLVPTFVSCILLGLDURSG 1138
 QY 1139 LNLISIVMILVDTVGFMALMDSYNVASHINIVSAVGSVBFVSHITRSFALSTKPTWL 1198
 DB 1139 LNLISIVMILVDTVGFMALMDSYNVASHINIVSAVGSVBFVSHITRSFALSTKPTWL 1198
 QY 1199 ERAKATISGSAVAVGANTNLPGLIIVGLAKAQLIQIFPRLNLLITLGLHGLVFL 1258
 DB 1199 ERAKATISGSAVAVGANTNLPGLIIVGLAKAQLIQIFPRLNLLITLGLHGLVFL 1258
 QY 1259 PVLLSYGPDVNPALMEQKRAEVAANVAVASCPNHPRSVSTADNLYVNHSPFGS1-KG 1317
 DB 1259 PVLLSYGPDVNPALMEQKRAEVAANVAVASCPNHPRSVSTADNLYVNHSPFGS1-KG 1317
 QY 1318 AGAISNPLPNNGRQF 1332
 DB 1318 AGAISNPLPNNGRQF 1332

RESULT 13

; US-10-646-301A-2
 ; Sequence 2, Application US/10646301A
 ; Publication No. US20040137467A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Altmann, Scott W
 ; APPLICANT: Wang, Luquan
 ; APPLICANT: Graziano, Michael
 ; APPLICANT: Murgolo, Nick
 ; TITLE OF INVENTION: NPC1L1 (NPC3) AND METHODS OF USE THEREOF
 ; FILE REFERENCE: J01603-K1-US
 ; CURRENT APPLICATION NUMBER: US/10/646,301A
 ; PRIOR FILING DATE: 2003-08-22
 ; PRIOR APPLICATION NUMBER: 60/397,442
 ; PRIOR FILING DATE: 2002-07-19
 ; PRIOR APPLICATION NUMBER: 10/621,758
 ; NUMBER OF SEQ ID NOS: 50
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 2
 ; LENGTH: 1331
 ; TYPE: PR1
 ; ORGANISM: Rattus sp.
 ; US-10-646-301A-2

Query Match 78.5%; Score 5421.5; DB 4; Length 1331;
 Best Local Similarity 77.8%; Pred. No. 0;
 Matches 1039; Conservative 120; Mismatches 169; Indels 7; Gaps 6;

QY 1 MAEAGRGWLLMALLRLAOSBPTTTHQPGYCAFYDECGKNPBLSGSLMTLSNVSCLSN 60
 DB 1 MAEAGRGWLLMALLRLAOSBPTTTHQPGYCAFYDECGKNPBLSGSLMTLSNVSCLSN 59
 QY 61 TPAARKITGDHLLILOKICPRLYNGPNTQ-ACCSAKOIVSLBASISITKALLTRCPACSDN 119
 DB 60 TPAARKITGDHLLILOKICPRLYNGPNTTACCSTKOLLSLESSWSITKALLTRCPACSDN 119

Qy	120	FVNHLCHNTCSBNOSLFTINVTNRVAOLGAGOLP	AVAVAYEAPFIOHSAFOSYSCSRVRVPA	179
Db	120	FVSLHCHNTCSBDOSLFTINVTNRVVERGEBPEP	VAAYEAPFORSFAEKVASCSCOVRI	179
Qy	180	AATLAVGMCVYSALCNAORMLNFQDGTNGLA	PLITHTHLEBPQAVSGSIOPLNEG	239
Db	180	AASLAVSGMCGVYSALCNAQMWLNFQDGTNGLA	PLITHTHLEBPQALPDGQIOPLNGK	239
Qy	240	VARCNESGQDVATSCSCODCAASC	PAIRAPALDSTFYLQMGPSLVIIILCSFAVAVT	299
Db	240	IAPCNESGQDSDAVSCSCODCAASC	VIIPPEALARSFTMGMPGMLALIIIFTA	299
Qy	300	ILLVGFVAPAPADSKRMVDPKKGTSLS	DKLSFSFTHLLIGFPQCGWTVASWPLTIVLS	359
Db	300	AVLVALRVASNRNKRKAGQEPAPFLPKHKL	SPHTILIGREFQWGTVAWMPPLTVLALS	359
Qy	360	VIPVVALAAGLVFTELTTPVEYELMSPNSO	ARSEAFHDOHFGPFPFRNOVILTA	419
Db	360	FIVIALAAGLFTIELTTPVELMSPAPSQ	ARKKESPFDEHGPFPFRNOQIVTA	419
Qy	420	YRVSLILGPKNFSIILDLILLELLEOERL	RHLQVMSPEAKORINISLODICVAP	479
Db	420	YKYSBILGSKNFSIILDLIFELLELEOERL	RHLQVMSPEAKORINISLODICVAP	479
Qy	480	TSLYDCCINSILQYQNNRTLLLTANQTL	MQOTSQVDMKDHFLYCANAPLTFKDG	539
Db	480	TSLSDDCVNSILQYQNNRTLLMTANQTL	MQOTSLVDMKHFLYCANAPLTFKDG	539
Qy	540	LSCMADYAPVPPFLATIGYKGYKDYSE	BALIMTSLNNTYPAQDPRLOAKLME	599
Db	540	LSMADYAPVPPFLAVAGYOGTDSE	BEALITFSLNNYPADPFRMAQALME	599
Qy	600	MRAFORAAGMOVFTFAERSLJEDIN	RTAEDDILIPASTYIVFLYLSL	659
Db	600	MESFORNTSDKQVAFSARSLJEDIN	RTITQDLPVFAVSITIVELYLSL	659
Qy	660	VMDVSKATLGLGVAVVLGA	VAAMAFPSYLGIRBSLVILQVPELV	719
Db	660	VAVESKATLGLGVAIVLGA	VLAAMGFSTYLGVPESVITIQVPELV	719
Qy	720	EYORLPRRPPGEBREVIHIGALGR	APSMULCSLSAICFPYLGALTPMP	779
Db	720	EYORLPRMPGEBREAHIGRTLG	SVAPSMULCSLSAICFPYLGALTP	779
Qy	780	VILDELOMSAFVALSLSDSKROE	ASRLDYVCCVAPQGLPRPGQSE	839
Db	780	IILDELOMTAFVALSLSDSKROE	SRPDLCCSTFRKLPPKEXEGL	839
Qy	840	LLHMTTRGVVLLFLFALFGVSL	YSMCHISVGLDQELALPKDSYL	899
Db	840	LLHMTTRVVMVLLFLFALFGAN	ULYKMCNINVGIDQELALPKDSYL	899
Qy	900	VYFVTTTIGYNFSSEAGNNAICSS	AGCNNSFTOKIQVATEPPEOSY	959
Db	900	VYFVTTSGFNFSSEAGNNAICSS	AGCKSFFSLQKIQVASEPPEOSY	959
Qy	960	WLTP-SSCCRLYISGPNCKDKCP	SPVYVNSLNCUKNOMSTMG	1019
Db	960	WLTPSSSCCRLYIRKPHKDE	PCPSTDTFNCCLKMCNMTL	1019
Qy	1019	RPNICPPGGLAAVSTVNLTS	DGQVLAASFMAHYKPKNSODY	1079
Db	1020	PPNIRCPGGLAAVSTVNLSS	DQVIAISQPMAYHKPLRNSOD	1079
Qy	1079	DLRKVPGTDPAEEVFPYTTITNV	FYEQYITLIPBGLFMLSLC	1139
Db	1080	DLRKVPGTDPMPEVFPYTTISNV	FYQYITLIPBGLFTALCP	1139
Qy	1139	LLNLISLIMLVDTYVGMALMDIS	TNAVSLINLVASGMSVEFVSH	1199
Db	1140	LLNLISIMILVDITIGLAWGIS	TNAVSLINLVAVGMSVEFVSH	1199

```

QY      1199 EBAEATISMGSAFAGVAMNTNPGIIIVGLAKAQLOIPEPFRNLITLLGLHGVFL 1258
         |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      1200 ERADDAIVFMGSAFAGVAMNTNPGIILLGPANQMLQIIEFFRNLITLLGLHGVFL 1259
QY      1259 PVLSYVGDPDVPNALALAEOKRAEEVAVAVMVASCPNHPSRVSTADNTIYNHSPFGSI-KG 1317
         |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      1260 PVLSTYLGPDNQALNVGEKLASBA-AVAPBPSCFPQIPSPADADAN--VNYGFAPBLAHG 1316
QY      1318 AGAISNFLPNNGROF 1332
         |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      1317 ANAARSLPKSQDKF 1331

RESULT 14
US-10-736-769-2
; Sequence 2, Application US/10736769
; Publication No. US20040161838A1
; GENERAL INFORMATION:
; APPLICANT: Altman, Scott W
; APPLICANT: Wang, Luquan
; APPLICANT: Graziano, Michael
; APPLICANT: Murgolo, Nick
; TITLE OF INVENTION: NPC1 (NPC3) AND METHODS OF USE THEREOF
; FILE REFERENCE: JBO1603-K3-US
; CURRENT FILING DATE: 2003-12-16
; PRIOR APPLICATION NUMBER: 60/3397,442
; PRIOR FILING DATE: 2002-07-19
; PRIOR APPLICATION NUMBER: 10/621,758
; PRIOR FILING DATE: 2003-07-17/646,301
; PRIOR APPLICATION NUMBER: 10/646,301
; PRIOR FILING DATE: 2003-08-22
; PRIOR APPLICATION NUMBER: 10/663,208
; PRIOR FILING DATE: 2003-09-16
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 1331
; TYPE: prt
; ORGANISM: Rattus sp.
US-10-736-769-2

Query Match          78.5%; Score 5421.5; DB 4; Length 1331;
Best Local Similarity 77.8%; Pred. No. 0;
Matches 1039; Conservative 120; Mismatches 169; Indels 7; Gaps 6;

QY      1 MAEAGRGWMLMALILRLAOSFPTTTHQPYCAFYBECGNPLSGSIMLTVSVCISN 60
DB      1 MAAAMWL-GWMLMALILLSAQBELVTTPKHEAVCTTFYEBCCGNPLSGTISLVSVCISN 59
QY      61 TPARKITSDDLILLOKICPRLTYTGPNTO-ACCASAKOLVSEBASISITKALLTRPCACSDN 119
         |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      60 TPAHVHTGELHALQRICPRLYNGPNNTTFACCSIKQOLLSSLESSMSITKALLTRPCACSDN 119
QY      120 FVNLIHCNTCSPNQSLFINTVTRVAQLGAGOLPAVVAYEAFAFOHSFAEOSYDCSRVAPA 179
         |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      120 FVSLJHCNTCSPDQSLFINTVTRVERGAGEBPVAVAYEAFAFORGAFAEKAVESSQVRIPA 179
QY      180 AATLAATGTCGYVGSALCNMQRMTNFCQGDGTNGALAPLDITTHLLEPQAVOGSGIQPLNEG 239
         |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      180 AASIAVGSIMCVYSALCANMRMTNFQGDGTNGALAPLDITFHLLPEQALPDGIQPLNGK 239
QY      240 VARCNESGGDDVAATCSCODCAASCPTAJARPOLDSTFYLGOMPSLVLIITLCGVFAVYT 299
         |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      240 IAPCNESGGDDSAVCSCODCAASCPTVIPPEBALRPSTYMGMPWMLAIITFTAFVULS 299
QY      300 ILVGFVAPARDXSKNWDPKKGTSLSDKLSPSTHTLLGQFQGMGTWVASWPILTIVLS 359
         |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      300 AVLVRIARVSVNRNNKXKAGEPEAKLPKHKLSPHTLTIGRFQVMWGTRVNASWPILTIALS 359
QY      360 VIPPVALAAGLVETELTDPEVLWSAPNSQARSKAFFHDQFGPPFRFTNOVIILAPNRS 419
         |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      360 FIIVATLAAGLTIELTDPVELWLSAPKSQARKSKSFDEHFEGFFRNQIFVTAIRNSS 419

```



```
Db      783 DFLQMTATFVALSLDSKQZASRPDYVCCFSSRNLPPEKQKEGLLCCFRKIYTPFLH 842
Qy      843 WITRGVLLFLFALFVSVLSYMCNHSVGLDOELALPKDSYLDYFPLNRYFEVGAPYF 902
Db      843 RIRPVLVLLFLVLPFANILYMCNISVGLDQDALPKDSYLDYFPLNRYLEVGPYF 902
Qy      903 VTTLLGNFSSEAGMNAICSSAGCNFSFTQKIQYATEPPEOSTYLAIPASSWDDFTDMLT 962
Db      903 DTSSGYNFSTEAGMNAICSSAGCESFTQKIQYASEFPNOSYVAIAASSWDDFTDMLT 962
Qy      963 P-SSCCRLYISGPNKXKPCPSNTNSLCKNCMSITMGSVRPSVEQFHXYLPWFLNDRPN 1021
Db      963 PSSSCCRITRGRHNDKDEPCPSDTSFNCKNCNRTLGPVRPTTEQFHXYLPWFLNDRPN 1022
Qy      1022 IKCPKGLAAYSTSVNLTDGVLASRFMAVHKPLKNSODYTEALRAARELANITADLR 1081
Db      1023 IKCPKGLAAYRTSVNLSSDGOIIASQFMAVHKPLKNSODFTALPASRLAANITAEIR 1082
Qy      1082 KVPGTDPAFEPPTITNVFYEQYTLTLPBGLFMLSCLVPTFAVSCLLGLDLRSGLN 1141
Db      1083 KVPGTDPNFEVFPYITISNVFYQOYTLVLPBGIPTTLALCFVPFVVCYLLGLDIRSGILN 1142
Qy      1142 LLSIVMLVDTYGFMAIMDISYNAVSLINLVSAVGSVEFVSHITRSFAISTKPTLERA 1201
Db      1143 LLSIIMILVDTIGLMAVWGISYNAVSLINLVTAVGMSVEFVSHITRSFAVSTKPTLERA 1202
Qy      1202 KEATISMSGSAVFAVAMTNLPGLVGLAKAQLIQIFPRLNLLITLLGLHGLVFLPYI 1261
Db      1203 KDATIFMSGSAVFAVAMTNFPGLILIGFAQAOIQIFPRLNLLITLLGLHGLVFLPYV 1262
Qy      1262 LSYVGPDVNPALALEQKRAEBAVAAMVASCPNHPRSVSTADNIYVNHSPFEGS-IKGAGA 1320
Db      1263 LSYLGPDVNQALVLEBKATEA-AMVSEBSCPOYPPADANTSDDYVNGFNPFIPEINA 1321
Qy      1321 ISNFLPNNGROF 1332
Db      1322 ASSSLPKSDQKF 1333
```

Search completed: March 22, 2006, 23:10:54
Job time : 194 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2006, 22:47:02 ; Search time 201 Seconds

(without alignments)
2911.704 Million cell updates/sec

Title: US-10-736-769-4

Sequence: 1 MAAAGRGWLMALTLRLAQ.....GSIKAGATSNFPPNNGRQF 1332

Scoring table:

BLOSUM62

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database:

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6909	100.0	1332	8	ADJ27297 Human NPC
2	6909	100.0	1332	9	ADY60740 Human NPC
3	6909	100.0	1332	9	ADY60740 Human NPC
4	6896	99.8	1332	4	AAW79169 Human pro
5	6896	99.8	1332	4	AAW79169 Human pro
6	6872.5	99.5	1359	8	ADJ27337 Human NPC
7	6872.5	99.5	1359	8	ADJ27337 Human NPC
8	6872.5	99.5	1359	8	ADY60780 Human NPC
9	6872.5	99.5	1359	8	ADY60780 Human NPC
10	6536	94.6	1344	4	ABG22693 Novel hum
11	5421.5	78.5	1331	8	ADJ27295 Rat NPCIL
12	5421.5	78.5	1331	8	ADY60738 Rat Niema
13	5421.5	78.5	1331	9	ABE93568 Rat NPCIL
14	5407	78.3	1333	8	ADJ27305 Mouse NPC
15	5407	78.3	1333	8	ADY60748 Mouse NPC
16	5407	78.3	1333	8	ADY60748 Mouse NPC
17	4466	64.6	982	4	ABG22691 Novel hum
18	2402.5	34.8	1278	2	AAW84445 Human NPC
19	2402.5	34.8	1278	2	AAW84445 Human NPC
20	2402.5	34.8	1278	8	ADQ39879 Human myo
21	2402.5	34.8	1278	8	ADQ39879 Human myo
22	2385	34.5	1319	2	AAW84446 Mouse NPC
23	2262	32.7	1287	4	ABE61737 Drosophil
24	2262	32.7	1287	8	ADY96670 Drosophil

25	1864.5	27.0	1223	4	ABE58629 Drosophil
26	1329	19.2	1170	2	AAW84447 Yeast NPC
27	1078	15.6	1296	2	AAW84448 Caenorhab
28	963	13.9	229	4	AAW80153 Human pro
29	963	13.9	229	4	AAW80152 Human pro
30	942	13.6	194	4	AAW5637 SSD domai
31	651.5	9.4	1447	2	AAW75375 Human pat
32	651.5	9.4	1447	2	AAW52200 Human pat
33	651.5	9.4	1447	2	AAW72969 Human pat
34	651.5	9.4	1447	4	AAW67163 Human pat
35	651.5	9.4	1447	5	AAW19830 Human pat
36	651.5	9.4	1447	5	ABJ10931 TRC8 rela
37	651.5	9.4	1447	5	AAW79571 Human pat
38	651.5	9.4	1447	7	AAW62275 Human pat
39	651.5	9.4	1447	7	ADY62678 Human pro
40	651.5	9.4	1447	7	ADY62678 Human pro
41	651.5	9.4	1447	7	ADY62678 Human pro
42	651.5	9.4	1447	8	ADY62678 Human pro
43	646.5	9.4	1434	2	AAW52199 Human pat
44	646.5	9.4	1434	2	AAW52199 Human pat
45	646.5	9.4	1434	4	AAW67159 Murine pa

ALIGNMENTS

RESULT 1	ADJ27297 standard; protein; 1332 AA.
ID	ADJ27297
XX	ADJ27297;
AC	ADJ27297;
XX	20-MAY-2004 (first entry)
DT	XX
XX	Human NPCIL.
XX	Niemann-Pick disease; type C1; gene-like 1; NPCIL; trans-golgi network;
KW	plasma membrane; transport signal; promoter;
KW	sterol regulated element binding protein 1; SREBP1;
KW	binding consensus sequence; transmembrane domain; sterol-sensing domain;
KW	SSD; cholesterol; NPCIL; receptor; Niemann-Pick C1 disease; intestinal;
KW	cholesterol absorption; serum cholesterol; hyperlipidaemia;
KW	atherosclerosis; coronary heart disease; stroke; arteriosclerosis.
XX	Homo sapiens.
OS	W0204009772-A2.
XX	29-JAN-2004.
XX	17-JUL-2003; 2003WO-US022467.
XX	19-JUL-2002; 2002US-0397442P.
XX	(SCHE) SCHERING CORP.
PA	Altmann SW, Murgolo NJ, Wang LQ, Graziano MP;
XX	WPI; 2004-132945/13.
XX	N-PSDB; ADJ27296.
DR	GENBANK; AF192522.
XX	New Niemann-Pick disease, type C1, gene-like 1 polypeptide, for detecting
PT	or identifying antagonists of NPCIL for inhibiting intestinal
PT	cholesterol absorption in a subject, or for treating elevated serum
PT	cholesterol or stroke.
XX	Claim 11; SEQ ID NO 4; 125bp; English.
PS	This sequence represents a Niemann-Pick disease, type C1, gene-like 1
CC	(NPCIL) polypeptide. NPCIL is an N-glycosylated protein which contains
CC	a motif ADJ27331 which acts as a trans-golgi network to plasma membrane
CC	transport signal, and which exhibits limited tissue distribution and

CC gastrointestinal abundance. The human NPC1L1 promoter sequence contains a
 CC sterol regulated element binding protein 1 (SREBP1) binding consensus
 CC sequence. NPC1L1 has 13 transmembrane spanning segments and a sterol-
 CC sensing domain (SSD) which is involved in sensing cholesterol levels,
 CC possibly by a mechanism which involves direct cholesterol bonding. NPC1L1
 CC has 42% amino acid sequence homology to human NPC1, a receptor
 CC responsible for Niemann-Pick C1 disease. The NPC1L1 polypeptides or
 CC polynucleotides are useful for detecting or identifying antagonists of
 CC NPC1L1, which can be used for inhibiting intestinal cholesterol
 CC absorption in a subject, or for treating medical conditions including
 CC elevated serum cholesterol, hyperlipidemia, atherosclerosis, coronary
 CC heart disease, stroke or arteriosclerosis.

XX Sequence 1332 AA;

Query Match 100.0%; Score 6909; DB 8; Length 1332;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1332; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAEAGLRGWLMLLRLAQSERYTTHQPGYCAFYDECGKNEPESGSLMTLSNVGCLSN 60
 DB 1 MEAGLRGWLMLLRLAQSERYTTHQPGYCAFYDECGKNEPESGSLMTLSNVGCLSN 60
 QY 61 TPARKITGTHLLLOKICRPLTYGPTNQACCSAKOVLSEASITKALLTRCPACSDNF 120
 DB 61 TPARKITGTHLLLOKICRPLTYGPTNQACCSAKOVLSEASITKALLTRCPACSDNF 120
 QY 121 VNLHCNTPSPNOSTLINTVAVQAGOLPAVVAEAFYQHSFAQSYDSCRAVEPA 180
 DB 121 VNLHCNTPSPNOSTLINTVAVQAGOLPAVVAEAFYQHSFAQSYDSCRAVEPA 180
 QY 181 ATLAVGTCGVYGSALCNARWLNFOGDTGNGLAPLDITFHLLPEQAVSGIOPINEGV 240
 DB 181 ATLAVGTCGVYGSALCNARWLNFOGDTGNGLAPLDITFHLLPEQAVSGIOPINEGV 240
 QY 241 AECNESQGDVATCSCODCAASCPAARPAOLDSTYVLGOMPSVLIIILCSVFAVVT 300
 DB 241 AECNESQGDVATCSCODCAASCPAARPAOLDSTYVLGOMPSVLIIILCSVFAVVT 300
 QY 301 LLVGFVAVPARDKSKWVDPKKGTSLSDKLSFTHTLLGPFQGMGTWVASMPLTIIVL 360
 DB 301 LLVGFVAVPARDKSKWVDPKKGTSLSDKLSFTHTLLGPFQGMGTWVASMPLTIIVL 360
 QY 361 IPVALAAGLVFTELTTDPVELMWSAPNSQARSEKAFHDGFPFRNQVILTA 420
 DB 361 IPVALAAGLVFTELTTDPVELMWSAPNSQARSEKAFHDGFPFRNQVILTA 420
 QY 421 RYDSLLGPRNFGIIDLLELLELLELLELLELLELLELLELLELLELLELLELLE 480
 DB 421 RYDSLLGPRNFGIIDLLELLELLELLELLELLELLELLELLELLELLELLE 480
 QY 481 SLVDCINSILQYFQNNRTLLLTANQTLMGQTSQVDMKDFLYCANAPLTFADGTAL 540
 DB 481 SLVDCINSILQYFQNNRTLLLTANQTLMGQTSQVDMKDFLYCANAPLTFADGTAL 540
 QY 541 SCWADYGAVPFPFLAIGYKXGKDYSEAEALIMTFSINNYPAQDPRLAQA 600
 DB 541 SCWADYGAVPFPFLAIGYKXGKDYSEAEALIMTFSINNYPAQDPRLAQA 600
 QY 601 RAFORMAGMFQVTFPERSLEDEINRTAEDPIPATSYIVFYITSLAGSYSSMSRY 660
 DB 601 RAFORMAGMFQVTFPERSLEDEINRTAEDPIPATSYIVFYITSLAGSYSSMSRY 660
 QY 661 MYDSKATLGLGAVAVLGAVMAMGFFSYLGRSSLYIIOVPELVLSVAGADNIF 720
 DB 661 MYDSKATLGLGAVAVLGAVMAMGFFSYLGRSSLYIIOVPELVLSVAGADNIF 720
 QY 721 YORLPRRPGEPREVIHGRALGRVAPSWLCSLEAICFFLGALTMPPAVTFALTGL 780
 DB 721 YORLPRRPGEPREVIHGRALGRVAPSWLCSLEAICFFLGALTMPPAVTFALTGL 780
 QY 781 IIDFLLQMSAFVALLSLDSKROEASRLDVCCYKPOELPPPGGEGILLGFPQKAYAP 840
 DB 781 IIDFLLQMSAFVALLSLDSKROEASRLDVCCYKPOELPPPGGEGILLGFPQKAYAP 840

DB 781 IIDFLLQMSAFVALLSLDSKROEASRLDVCCYKPOELPPPGGEGILLGFPQKAYAP 840
 QY 841 LHMITRGVLLFLALFGVSLVSMCHISVGLDOELAPKDSYLLDYFLNRYFEVGA 900
 DB 841 LHMITRGVLLFLALFGVSLVSMCHISVGLDOELAPKDSYLLDYFLNRYFEVGA 900
 QY 901 YPFTTIGYVSSBAGNMAICSSAGCNSBFQKIQTATPEPPOSYLAIPASSVVD 960
 DB 901 YPFTTIGYVSSBAGNMAICSSAGCNSBFQKIQTATPEPPOSYLAIPASSVVD 960
 QY 961 LTPSSCCRLYISGPNKDKCPSTVNSLANCKNCKMSITMGSVPRSVBQFHKYLP 1020
 DB 961 LTPSSCCRLYISGPNKDKCPSTVNSLANCKNCKMSITMGSVPRSVBQFHKYLP 1020
 QY 1021 NIKCPKGLAAYSTSVNLTSQVLAARMAYHKPKNSQDYTEALRAARELAANTAD 1080
 DB 1021 NIKCPKGLAAYSTSVNLTSQVLAARMAYHKPKNSQDYTEALRAARELAANTAD 1080
 QY 1081 RKVPGTDPAPFVFPYITTVFBOYLTIPPEGLFMISLCLVTFPAVSCILLGLD 1140
 DB 1081 RKVPGTDPAPFVFPYITTVFBOYLTIPPEGLFMISLCLVTFPAVSCILLGLD 1140
 QY 1141 NLLSTVMIIVDTYGFMAIMDISYNAVSLINLVSAVMSYEFVSHITRSPAI 1200
 DB 1141 NLLSTVMIIVDTYGFMAIMDISYNAVSLINLVSAVMSYEFVSHITRSPAI 1200
 QY 1201 AKERATISMSAVFAGVAMTNLPGLIVLGLAKQILQIPFFRLNLLITLLGLH 1260
 DB 1201 AKERATISMSAVFAGVAMTNLPGLIVLGLAKQILQIPFFRLNLLITLLGLH 1260
 QY 1261 ILSYGPDPVNPALALEQKAEBAVAAVMVASCNPHSRVSTADNIVNHSFES 1320
 DB 1261 ILSYGPDPVNPALALEQKAEBAVAAVMVASCNPHSRVSTADNIVNHSFES 1320
 QY 1321 ISNFLPNNGROF 1332
 DB 1321 ISNFLPNNGROF 1332

RESULT 2
 ID ADY60740 standard; protein; 1332 AA.
 AC ADY60740;
 DT 19-MAY-2005 (first entry)
 XX XX
 DE Human Niemann-Pick C1-like protein 1 (NPC1L1) protein seq ID 4.
 XX XX
 KW Niemann-Pick C1-like protein 1; Antilipemic; Antiarteriosclerotic;
 KW Cardiac; Vascular; Cerebroprotective; Cholesterol; hyperlipidemia;
 KW metabolic disorder; atherosclerosis; cardiovascular disease;
 KW coronary artery disease; cerebrovascular ischemia; neurological disease;
 KW arteriosclerosis; transgenic animal.
 XX XX
 OS Homo sapiens.
 OS XX
 PN MO2005015988-A1.
 XX XX
 PD 24-FEB-2005.
 XX XX
 PF 16-DEC-2003; 2003MO-US040113.
 XX XX
 PR 17-JUL-2003; 2003US-00621758.
 PR 22-AUG-2003; 2003US-00646301.
 PR 16-SEP-2003; 2003US-00663208.
 XX XX
 PA (SCHE) SCHERING CORP.
 XX XX
 PI Altman SW, Murgolo NJ, Wang LQ, Graziano MP;
 XX XX
 DR WPI; 2005-284403/29.
 DR N-PSDB; ADY60739, ADY60787.

XX Novel isolated NPC1L1 polypeptide, useful for identifying antagonist of
 PT NPC1L1 that inhibits NPC1L1 mediated sterol or 5 alpha-sterol uptake, for
 PT treating hyperlipidemia or atherosclerosis in mouse, rat or human.
 XX Claim 30, SEQ ID NO 4, 146pp, English.

XX The invention relates to an isolated Niemann-Pick C1-like protein 1
 CC (NPC1L1) polypeptide comprising 42 or more contiguous amino acids from an
 CC amino acid sequence appearing as ADY60738, ADY60740 or ADY60748 (Rat,
 CC human and mouse NPC1L1). Also included are an isolated polynucleotide
 CC encoding NPC1L1 (appearing as ADY60737, ADY60739 or ADY60747), a
 CC recombinant vector comprising the polynucleotide, a host cell comprising
 CC the vector, an isolated antibody which specifically binds to NPC1L1 (or
 CC to a rat NPC1L1 peptide sequence appearing as ADY60775-ADY60778),
 CC preparation of NPC1L1, a mutant transgenic mouse comprising a homozygous
 CC mutation of endogenous chromosomal NPC1L1 (where the mouse does not
 CC produce any functional NPC1L1 protein), an offspring or progeny of the
 CC mouse having inherited a mutated NPC1L1 allele of the mouse, screening a
 CC sample for an intestinal sterol or Salpha-sterol absorption antagonist,
 CC inhibiting NPC1L1 mediated sterol or Salpha-sterol uptake in a subject, a
 CC kit (comprising ezetimibe or substituted azetidione in a pharmaceutical
 CC dosage form, and information indicating that NPC1L1 is a target of
 CC ezetimibe or substituted azetidione), decreasing the level of intestinal
 CC sterol or Salpha-sterol absorption in a subject (involving reducing the
 CC level of expression of NPC1L1 in the subject), identifying an antagonist
 CC of NPC1L1 and an isolated mammalian cell which lacks a gene encoding a
 CC functional NPC1L1 protein. NPC1L1 is useful for identifying an antagonist
 CC of NPC1L1 which is useful for inhibiting or decreasing the level of
 CC NPC1L1 mediated sterol or Salpha-sterol uptake in a subject such as
 CC mouse, rat or human. The level of expression of NPC1L1 in the subject is
 CC reduced by mutating NPC1L1 in the subject. The NPC1L1 knockout mouse is
 CC useful for screening a sample for intestinal sterol or Salpha-sterol
 CC absorption antagonist. The NPC1L1 antagonists identified are useful for
 CC treating medical conditions caused or mediated by NPC1L1, e.g.,
 CC hyperlipidemia, atherosclerosis, coronary heart diseases, stroke or
 CC arteriosclerosis. The present sequence represents a mammalian NPC1L1
 CC protein.

XX Sequence 1332 AA:

Query Match 100.0%; Score 6909; DB 9; Length 1332;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1332; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 MAEAGRGMLLWALLRLAOSPTTTHQPGYCAFYDECGKNPBLSGSLMTLSNVCLSN 60
 QY 1 TPAKRTGDHLILLOKICPRLYTGPNTQACCSAKOLVSLASISITKALTTRCPACSDNF 120
 DB 61 TPAKRTGDHLILLOKICPRLYTGPNTQACCSAKOLVSLASISITKALTTRCPACSDNF 120
 QY 121 VNIHCHTCSPNOSLFINTRVAQLGAGQLPAVVAEAFYOHSEAEOSYDSCSRVRPAA 180
 DB 121 VNIHCHTCSPNOSLFINTRVAQLGAGQLPAVVAEAFYOHSEAEOSYDSCSRVRPAA 180
 QY 121 VNIHCHTCSPNOSLFINTRVAQLGAGQLPAVVAEAFYOHSEAEOSYDSCSRVRPAA 180
 DB 121 VNIHCHTCSPNOSLFINTRVAQLGAGQLPAVVAEAFYOHSEAEOSYDSCSRVRPAA 180
 QY 181 ATTAGVTCGVYSALCNAQRWLNFGQDTGNGLAPLDITFHLLPEGQAVSGIOLNEGV 240
 DB 181 ATTAGVTCGVYSALCNAQRWLNFGQDTGNGLAPLDITFHLLPEGQAVSGIOLNEGV 240
 QY 241 ARCNESGDDVAVTSCDDCAASCPAARPAOLDSTFVLGMPGSLVLIILGCVFAVVTI 300
 DB 241 ARCNESGDDVAVTSCDDCAASCPAARPAOLDSTFVLGMPGSLVLIILGCVFAVVTI 300
 QY 301 LTVGFRAVAPARDISKWDPKKGTSLSDKLSFSTHTLLGQFFQGWGTWVWASMPLTIIIVLSV 360
 DB 301 LTVGFRAVAPARDISKWDPKKGTSLSDKLSFSTHTLLGQFFQGWGTWVWASMPLTIIIVLSV 360
 QY 361 IPVVALAAGLVFTLTDVPELWSPNSQARSEKAFHDFGPFRRNOVILTRPNSSY 420
 DB 361 IPVVALAAGLVFTLTDVPELWSPNSQARSEKAFHDFGPFRRNOVILTRPNSSY 420

QY 421 RYDSLLGPKNPSGIIIDDLLELELQERLRLHQLQVSPBAQNISLQDICTAPLNPDT 480
 DB 421 RYDSLLGPKNPSGIIIDDLLELELQERLRLHQLQVSPBAQNISLQDICTAPLNPDT 480
 QY 481 SLVDCCLNSLLQFQNNRTLLLTANQTLMGQTSQVMDQHPFLYCANAPLTFPDGTALAL 540
 DB 481 SLVDCCLNSLLQFQNNRTLLLTANQTLMGQTSQVMDQHPFLYCANAPLTFPDGTALAL 540
 QY 541 SCWADGAVFPFPLAIGYKGDYSEAEALIMTFSANNYPAGDPRLAQAWEAELEEM 600
 DB 541 SCWADGAVFPFPLAIGYKGDYSEAEALIMTFSANNYPAGDPRLAQAWEAELEEM 600
 QY 601 RAFQRMAGMFOVTFPAERSLDEIRRTAEDLPITATSIYVFLYISLALGSSWSRV 660
 DB 601 RAFQRMAGMFOVTFPAERSLDEIRRTAEDLPITATSIYVFLYISLALGSSWSRV 660
 QY 661 MDSKATTLGCGAVVILGAVMAAMGFSTLGRSSVLIQVPELVLSGADNIFLTVLE 720
 DB 661 MDSKATTLGCGAVVILGAVMAAMGFSTLGRSSVLIQVPELVLSGADNIFLTVLE 720
 QY 721 YORLPRRPGPREVHIGRALGRVAPSMILCSLSEALCFPLGALTTPPAVTFALTGLAV 780
 DB 721 YORLPRRPGPREVHIGRALGRVAPSMILCSLSEALCFPLGALTTPPAVTFALTGLAV 780
 QY 781 IUDFLQMSAFVALSLDSKROEASRLDVCCVKGPELPPGSGILLGFQKAYAPFL 840
 DB 781 IUDFLQMSAFVALSLDSKROEASRLDVCCVKGPELPPGSGILLGFQKAYAPFL 840
 QY 841 LHWITGVVILLFLAFGLSVSMCHISVGLDQELAPKDSYLLDVFPLNRFVEVGAPV 900
 DB 841 LHWITGVVILLFLAFGLSVSMCHISVGLDQELAPKDSYLLDVFPLNRFVEVGAPV 900
 QY 901 YPFTTLGVNPSSEAGNNAICSSAGCNFSFTOKIOYATEPEBOSYAIIPASSWVDFIDW 960
 DB 901 YPFTTLGVNPSSEAGNNAICSSAGCNFSFTOKIOYATEPEBOSYAIIPASSWVDFIDW 960
 QY 961 LTPSSCCRLYISGPNKDKCPSTVNSLNCIKNCMSTMGSVRDSVEQFHXYLPWFLNDRP 1020
 DB 961 LTPSSCCRLYISGPNKDKCPSTVNSLNCIKNCMSTMGSVRDSVEQFHXYLPWFLNDRP 1020
 QY 1021 NIKCPKGLAAYSTVNLTSQGYLASRPMAYHKPKNSQDTYEAARAELEAANIITADL 1080
 DB 1021 NIKCPKGLAAYSTVNLTSQGYLASRPMAYHKPKNSQDTYEAARAELEAANIITADL 1080
 QY 1081 RKVPGTDPAFEPFYTITVFEQYLTILPEGLFMLSCLVPTFAVSCLLGLDLSGGL 1140
 DB 1081 RKVPGTDPAFEPFYTITVFEQYLTILPEGLFMLSCLVPTFAVSCLLGLDLSGGL 1140
 QY 1141 NLSIYVILVDITGFNALMDISYNAVSLINLVSAVQMSVEFVSHITRSFALSTKPTWLER 1200
 DB 1141 NLSIYVILVDITGFNALMDISYNAVSLINLVSAVQMSVEFVSHITRSFALSTKPTWLER 1200
 QY 1201 AKKATISMGSAVAGVAMTNLPGIIVLGLAKAQLIOIFFRRLNLTITLGLHLGVFLPV 1260
 DB 1201 AKKATISMGSAVAGVAMTNLPGIIVLGLAKAQLIOIFFRRLNLTITLGLHLGVFLPV 1260
 QY 1261 ILSYGPDPVNPALAEOKAEBAVAAVMAVASCNHSRSVSTADNIVNHSFEGSIGAGA 1320
 DB 1261 ILSYGPDPVNPALAEOKAEBAVAAVMAVASCNHSRSVSTADNIVNHSFEGSIGAGA 1320
 QY 1321 ISNPLPNNGROF 1332
 DB 1321 ISNPLPNNGROF 1332

RESULT 3
 AEB93570 ID AEB93570 standard; protein; 1332 AA.

XX AEB93570;

DT 06-OCT-2005 (first entry)

DE Human NPC1L1 protein.
KW Protein engineering; NPC1L1; antidiabetic; antiarteriosclerotic;
KW cardiant; vasotropic; hyperlipidemia; atherosclerosis; stroke;
KW arteriosclerosis; human.
XX Homo sapiens.
PN WO2005069900-A2.
XX 04-AUG-2005.
PD
PF 14-JAN-2005; 2005MO-US001469.
PR 16-JAN-2004; 2004US-0537341P.
PA (MERI) MERCK & CO INC.
PI Garcia-Calvo M;
XX WPI; 2005-564070/57.
DR N-PSDB; AEB93569.
XX
PT Identifying ligand of NPC1L1 for stimulating the activity of NPC1L1, by
PT contacting NPC1L1 with detectably labeled substituted 2-azetidinone
PT glucuronide and a candidate compound and determining if compound binds to
PT human NPC1L1.
PS Example 2; SEQ ID NO 4; 215pp; English.
XX
CC The invention relates to identifying a ligand of NPC1L1. The method
CC involves contacting human NPC1L1 with a detectably labeled substituted 2-
CC azetidinone glucuronide and a candidate compound and determining whether
CC the candidate compound binds to human NPC1L1. In identifying a ligand of
CC NPC1L1, the KD value is 100 nM or lower, preferably 50 nM, 20 nM, or 10
CC nM or lower. The detectably labeled substituted 2-azetidinone glucuronide
CC is 35 S-labeled compound 2, given in the specification. NPC1L1 ligands
CC are useful for stimulating or blocking the activity of NPC1L1, and for
CC treating conditions caused or mediated by NPC1L1. It is useful for
CC reducing the incidence of hyperlipidemia, atherosclerosis, coronary heart
CC disease, stroke, or arteriosclerosis. The present sequence represents a
CC human NPC1L1, a N-glycosylated protein.
XX
SQ Sequence 1332 AA;
Query Match 100.0%; Score 6909; DB 9; Length 1332;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1332; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 361 IPVVALAAGLVFTELTTDPVELMSAPNSQARSEKA FHOHGGPFRPTNOVILLTAPNRSY 420
DB 361 IPVVALAAGLVFTELTTDPVELMSAPNSQARSEKAFHDDHGGPFRPTNOVILLTAPNRSY 420
QY 421 RYDSLILGPKNSGSLDLDLLELELOERLHLQWSPQAQNTSLQDICYAPLPNDT 480
DB 421 RYDSLILGPKNSGSLDLDLLELELOERLHLQWSPQAQNTSLQDICYAPLPNDT 480
QY 481 SLYDCCINSLLQYFQNNRTLLLTANQTLMGQTSQVDWKDHFLYCANAPLTKDGTALAL 540
DB 481 SLYDCCINSLLQYFQNNRTLLLTANQTLMGQTSQVDWKDHFLYCANAPLTKDGTALAL 540
QY 541 SCMAVYGAAPVEPFLAIGYKGDYSEAEALIMTFSLNNYPADDPRLQAKLMEBAFLBEM 600
DB 541 SCMAVYGAAPVEPFLAIGYKGDYSEAEALIMTFSLNNYPADDPRLQAKLMEBAFLBEM 600
QY 601 RAFORRMAGMPQVTFEATERSLEDEINRTTAEPLIFATSYIIFLYISALGSYSWSRY 660
DB 601 RAFORRMAGMPQVTFEATERSLEDEINRTTAEPLIFATSYIIFLYISALGSYSWSRY 660
QY 661 WYDSKATLGLGVAVALGAVMAAMGFPSYLGIRSSLVILLQVVPFLVSVGADNIFLVELE 720
DB 661 WYDSKATLGLGVAVALGAVMAAMGFPSYLGIRSSLVILLQVVPFLVSVGADNIFLVELE 720
QY 721 YQRLRRRBEPREVHIGALGRVAPSMILCSISEALCFPLGALTMPANRTFALSGLA 780
DB 721 YQRLRRRBEPREVHIGALGRVAPSMILCSISEALCFPLGALTMPANRTFALSGLA 780
QY 781 IIDFLLQMSAFVALLSLDSKROEASRLDVCCVCKQELPPPaQGGELLLGFFOKAVAPPL 840
DB 781 IIDFLLQMSAFVALLSLDSKROEASRLDVCCVCKQELPPPaQGGELLLGFFOKAVAPPL 840
QY 841 LHMTRGVVLLFLALFGVSLYSWCHISVGLDQELAPKDSYLDLPLFLNRYFEVAVP 900
DB 841 LHMTRGVVLLFLALFGVSLYSWCHISVGLDQELAPKDSYLDLPLFLNRYFEVAVP 900
QY 901 YVTTLLGVNSESSEAGMNAICSSAGCNNSFPQKIYATEFPQSYLAPASSWDDFDWM 960
DB 901 YVTTLLGVNSESSEAGMNAICSSAGCNNSFPQKIYATEFPQSYLAPASSWDDFDWM 960
QY 961 LTPSSCCRLYISGPKDKFCPSTVNSLNCIAKNCISITGWSVAPSEOFHKYLPWFLNDRP 1020
DB 961 LTPSSCCRLYISGPKDKFCPSTVNSLNCIAKNCISITGWSVAPSEOFHKYLPWFLNDRP 1020
QY 1021 NIKCPKGLAAYSTSVNLTSQVLA SRPMAYHKPLKNSQDYTEALRAARELAANITADL 1080
DB 1021 NIKCPKGLAAYSTSVNLTSQVLA SRPMAYHKPLKNSQDYTEALRAARELAANITADL 1080
QY 1081 RRVPGTDAFEVFPYTTITNVFEQYLTLLPBGFLMILSLCLVPTFAVSCLLGLDRSGLL 1140
DB 1081 RRVPGTDAFEVFPYTTITNVFEQYLTLLPBGFLMILSLCLVPTFAVSCLLGLDRSGLL 1140
QY 1141 NLLSTYMLIVDTVGFMAIMDISYNAVSLINLVASAGMSVEFASHITRSFAISTKPTWLER 1200
DB 1141 NLLSTYMLIVDTVGFMAIMDISYNAVSLINLVASAGMSVEFASHITRSFAISTKPTWLER 1200
QY 1201 AKEATISMGSAVFAGVAMTNLPGLIIVLGLAKQOLIQIFFFRLLNLTLLGLLHGLVPLPV 1260
DB 1201 AKEATISMGSAVFAGVAMTNLPGLIIVLGLAKQOLIQIFFFRLLNLTLLGLLHGLVPLPV 1260
QY 1261 IISYVGPVNPALALEQKRAEBAVAAMVASCPNHPBSRVSTADNITYNHSFEGSIKGA 1320
DB 1261 IISYVGPVNPALALEQKRAEBAVAAMVASCPNHPBSRVSTADNITYNHSFEGSIKGA 1320
QY 1321 ISNPLPNNGRQF 1332
DB 1321 ISNPLPNNGRQF 1332
RESULT 4
AAM79169
ID AAM79169 standard; protein; 1332 AA.
XX

AC AAM79169;
XX 06-NOV-2001 (first entry)
XX
DE Human protein SEQ ID NO 1831.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorder; arthritis; inflammation.
OS Homo sapiens.
XX
XX MO200157190-A2.
XX
XX 09-AUG-2001.
XX
XX 05-FEB-2001; 2001WO-US004098.
XX
XX 03-FEB-2000; 2000US-00496914.
XX 27-APR-2000; 2000US-00560875.
XX 20-JUN-2000; 2000US-00598075.
XX 19-JUL-2000; 2000US-00620325.
XX 01-SEP-2000; 2000US-00634936.
XX 15-SEP-2000; 2000US-00663561.
XX 20-OCT-2000; 2000US-00693325.
XX 30-NOV-2000; 2000US-00728422.
XX
XX (HYSE-) HYSEO INC.
XX
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
XX Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
XX Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
XX WPI; 2001-476283/51.
XX N-PSDB; AAK52302.
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities, useful
XX in diagnosis and gene therapy.
XX
XX Claim 20; Page 4216-4219; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
XX encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activity/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukemia, nervous system disorders, arthritis and
XX inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
XX (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
XX sequence listing were missing at the time of publication
SQ Sequence 1332 AA;
Query Match 99.8%; Score 6896; DB 4; Length 1332;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1330; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 MAEAGLGLWLLMALLLRLAAGEPYTTIHOPGCAFYDECGNRPGLSGMLTSLNVSCLSN 60
DB 1 MAEAGLGLWLLMALLLRLAAGEPYTTIHOPGCAFYDECGNRPGLSGMLTSLNVSCLSN 60
QY 61 TPARKITGDHLLILLOKICPRLYTPNTQACCSAKQVLSLEASLSITRALLTRCPACSDNF 120
DB 61 TPARKITGDHLLILLOKICPRLYTPNTQACCSAKQVLSLEASLSITRALLTRCPACSDNF 120
QY 121 VNLHCNHTGSPNOSLFTNTRVAQLGAGOLPAVVAYAEAFYOHSAEASYSYSCSRVPPAA 180
DB 121 VNLHCNHTGSPNOSLFTNTRVAQLGAGOLPAVVAYAEAFYOHSAEASYSYSCSRVPPAA 180

QY 181 ATLAAGTMCVYSALCNAQRWLNFOGDTGNGLAPLDITFHLLEPGQAVSGSIGPLNEGV 240
DB 181 ATLAAGTMCVYSALCNAQRWLNFOGDTGNGLAPLDITFHLLEPGQAVSGSIGPLNEGV 240
QY 241 ARCNESQGDVAVACSCODCAASCPATARPQALDSTFVLGQMPGSVYLIIILCSVPANVTI 300
DB 241 ARCNESQGDVAVACSCODCAASCPATARPQALDSTFVLGQMPGSVYLIIILCSVPANVTI 300
QY 301 LLVGFRAVPARDKSKWVDPKKGTSLSDKLSFSTHTLLGQFQFGMGTVVASMPTLILVLSV 360
DB 301 LLVGFRAVPARDKSKWVDPKKGTSLSDKLSFSTHTLLGQFQFGMGTVVASMPTLILVLSV 360
QY 361 IPVVALAAGLVFTELTTDPVELWSAPNSQARSEKAFHDQFGPPFRITNOYLTPANRSSY 420
DB 361 IPVVALAAGLVFTELTTDPVELWSAPNSQARSEKAFHDQFGPPFRITNOYLTPANRSSY 420
QY 421 RYDLSLLGPNKFSGLIDDLLELLELLELLELLELLELLELLELLELLELLELLELLELLEL 480
DB 421 RYDLSLLGPNKFSGLIDDLLELLELLELLELLELLELLELLELLELLELLELLELLELLEL 480
QY 481 SLVDCCINSLLQYFQNNRTLLLTANQTLMGQTSQVDMKDFPYCANAPLTFDGTALAL 540
DB 481 SLVDCCINSLLQYFQNNRTLLLTANQTLMGQTSQVDMKDFPYCANAPLTFDGTALAL 540
QY 541 SCNADYGAVPFPLAIGYKGDYSEAEALIMTFSLNNYPADGPRLAQAALMEBAFLEEM 600
DB 541 SCNADYGAVPFPLAIGYKGDYSEAEALIMTFSLNNYPADGPRLAQAALMEBAFLEEM 600
QY 601 RAQORRAGMFOYTFPARESLDEINRTTADLPFATSTIYVFLYSLALGSYSMSRV 660
DB 601 RAQORRAGMFOYTFPARESLDEINRTTADLPFATSTIYVFLYSLALGSYSMSRV 660
QY 661 MDSKATLGLGVAVVLGAVMAAMGPFSSYLGIRSSLYILOVFPVLVSAGADNIFFLVLE 720
DB 661 MDSKATLGLGVAVVLGAVMAAMGPFSSYLGIRSSLYILOVFPVLVSAGADNIFFLVLE 720
QY 721 YORLPRRGPBREVHIGRALGRVAPSMWLCSLSEALICFGLALTMPAVRTFALTSGLAV 780
DB 721 YORLPRRGPBREVHIGRALGRVAPSMWLCSLSEALICFGLALTMPAVRTFALTSGLAV 780
QY 781 ILDFLLQMSAFVALLSLDSRQASRLDVCCYKPOELPPRGGEGLLGFPOKAVAPFL 840
DB 781 ILDFLLQMSAFVALLSLDSRQASRLDVCCYKPOELPPRGGEGLLGFPOKAVAPFL 840
QY 841 LHWITGVVLLFLALFGVSLVSMCHSVGLDDELALPKOSYLDYFLFLNRYEYGA 900
DB 841 LHWITGVVLLFLALFGVSLVSMCHSVGLDDELALPKOSYLDYFLFLNRYEYGA 900
QY 901 YFVTTIGYNSSEAGMNAICSSAGCNNSFTQKIQVATEPPEOSYLAIPASSWVDPEIDW 960
DB 901 YFVTTIGYNSSEAGMNAICSSAGCNNSFTQKIQVATEPPEOSYLAIPASSWVDPEIDW 960
QY 961 LTPSSCCRLYISGPNKXCPSTVNSLNCCKMSITMGSVRPSVBOFHKLTPWFLNDRP 1020
DB 961 LTPSSCCRLYISGPNKXCPSTVNSLNCCKMSITMGSVRPSVBOFHKLTPWFLNDRP 1020
QY 1021 NIKCPKGLAAYSTSVNLSDGOVLASRFMAHYKPKNSODYTEALPAARELANITADL 1080
DB 1021 NIKCPKGLAAYSTSVNLSDGOVLASRFMAHYKPKNSODYTEALPAARELANITADL 1080
QY 1081 RKVPGTDPAFEVPPYITTVFYBOYLLTLEBGFMLSLCLVPTFAVSCLLIGDLRSGLL 1140
DB 1081 RKVPGTDPAFEVPPYITTVFYBOYLLTLEBGFMLSLCLVPTFAVSCLLIGDLRSGLL 1140
QY 1141 NLSIWIILVDTYGFMALMDISYNAVSLINLVAGVMSVFEVGHITRSFALSTKPTWLER 1200
DB 1141 NLSIWIILVDTYGFMALMDISYNAVSLINLVAGVMSVFEVGHITRSFALSTKPTWLER 1200
QY 1201 AKKATISMSGSAVFGVAMTULPGILVYGLAKAOLIQIFFRMLNLTLLGLGLVFLPV 1260
DB 1201 AKKATISMSGSAVFGVAMTULPGILVYGLAKAOLIQIFFRMLNLTLLGLGLVFLPV 1260

Oy	1261	ILSYVGPDPVNPAALALEQRAEEAAVAAMVAVSCPNHPSRVSTADNITYNHSEFSITKGAGA	1320
Dd	1261	ILTSYGPDPVNPAALAEQRABEAAVAAMVAVSCPMPHSRVSTADNIYYNHSEFSGIKGAGA	1320
Oy	1321	ISNFLENNGRPF	1332
Dd	1321	ISNFLPNNGRPF	1332
 RESULT 5 AAG65638 ID AAG65638 standard; protein; 1332 AA. XX AC AAG65638; XX DT 07-JAN-2002 (first entry) XX DE SSD-containing SSP1 protein sequence.			
XX	SSD, steroid-sensing domain; human; liver; testis; brain; cancer;		
KW	nootropic; neuroprotective; antidiabetic; antiarteriosclerotic;		
XX	cycostatic; antilipemic; SSPL.		
OS	Homo sapiens.		
PN	WO200170974-A1.		
PD	27-SEP-2001.		
PE	22-MAR-2001; 2001WO-JP002279.		
PR	24-MAR-2000; 2000JP-00088595.		
PA	(TAKE) TAKEDA CHEM IND LTD.		
XX	Taniyama Y, Kita S, Komiya T;		
PI	WPI; 2001-611501/70.		
DR	N-PSDB; AAI66914.		
PT	New steroid-sensing domain-containing protein for diagnosing and screening		
XX	candidate compounds in drug development for diabetes, obesity, cancer,		
XX	arteriosclerosis, hyperlipidemia and neurodegenerative disorders.		
PS	Claim 7; Page 115-122; 171pp; Japanese.		
CC	The invention provides a novel SSD (steroid-sensing domain)-containing		
CC	protein. The protein originates from human liver, human testis or human		
CC	brain. The protein can be expressed by standard recombinant methodology.		
CC	The proteins, encoded DNAs and antibodies are useful in diagnosis and		
CC	screening candidate compounds in drug development for diabetes, obesity,		
CC	cancer, arteriosclerosis, hyperlipemia, neurodegenerative disorders such		
CC	as Alzheimer's disease and neural disorders. The present sequence		
CC	represents the SSPL protein which contains the SSD domain		
XX	Sequence 1332 AA;		
Oy	Query Match	99.8%; Score 6896; DB 4; Length 1332;	
Dd	Best Local Similarity	99.8%; Pred. No. 0;	
Oy	Matches 1330; Conservative	0; Mismatches 2; Indels 0; Gaps 0;	
Dd	1 MAEAGLRGMLLMALRLRLOSEPPTYTHIQPGCAFYDCGNKPELSGLMTLISNVSCLSN	60	
Oy	61 TPARKITGDHILLOKICPRLYTGENTQACCSAKOLVLEASLSITKALLTRCPACSDNF	120	
Dd	61 TPARKITGDHILLOKICPRLYTGENTQACCSAKOLVLEASLSITKALLTRCPACSDNF	120	
Oy	121 VNLGHNTCTSPNGSLFIIVNTRYAQLGAGQLPAVVAYEAFFYOHSPAEQSYDSCSRVPDAA	180	
Dd	121 VNLGHNTCTSPNGSLFIIVNTRYAQLGAGQLPAVVAYEAFFYOHSPAEQSYDSCSRVPDAA	180	

QY	181	ATLAVGMCYGGYSGALCNQARWLNFOGDDTNGLAPLDITTHLEBPGQAVSGIOPLNEGV	240
Db	181	ATLAVGMCYGGYSGALCNQARWLNFOGDDTNGLAPLDITTHLEBPGQAVSGIOPLNEGV	240
QY	241	ARCNEOGDDVATCSQDCDCAASCFAIARPALDSTFLGQMPGSLVLIILICSVFAVVTI	300
Db	241	ARCNEOGDDVATCSQDCDCAASCFAIARPALDSTFLGQMPGSLVLIILICSVFAVVTI	300
QY	301	LLVGFRAVAPADKSKMVDPKKGTSLDKLSFSTHTLLIGQFPQMGTVASWPLTILVLVS	360
Db	301	LLVGFRAVAPADKSKMVDPKKGTSLDKLSFSTHTLLIGQFPQMGTVASWPLTILVLVS	360
QY	361	IPVVALAAGVFEITLTDPEVLMSAPSOARSEKAFHDHGPFRFNNOVILTAPNRSSY	420
Db	361	IPVVALAAGVFEITLTDPEVLMSAPSOARSEKAFHDHGPFRFNNOVILTAPNRSSY	420
QY	421	RYDLSLLGPKNFSGLDLDLLELLELQERLRLHQVMSPEAQRNISIQDICIYAPLANDNT	480
Db	421	RYDLSLLGPKNFSGLDLDLLELLELQERLRLHQVMSPEAQRNISIQDICIYAPLANDNT	480
QY	481	SLYDCCINSLLQYFQNNRTLLLLTANQTLMGQTSQVDMKDHFLYCANAPLTFKXGTALAL	540
Db	481	SLYDCCINSLLQYFQNNRTLLLLTANQTLMGQTSQVDMKDHFLYCANAPLTFKXGTALAL	540
QY	541	SCMADYAPAPFPFLATIGGYKGYSEAEALIMFSLNNTYPAQDPRLOAKIMEBAFLEEM	600
Db	541	SCMADYAPAPFPFLATIGGYKGYSEAEALIMFSLNNTYPAQDPRLOAKIMEBAFLEEM	600
QY	601	RAFORBAMGQVYFTABERSLEDEINTTMBDLPFATSYIVFPLYSLALGSYSSMSRV	660
Db	601	RAFORBAMGQVYFTABERSLEDEINTTMBDLPFATSYIVFPLYSLALGSYSSMSRV	660
QY	661	MVDSKATLGLGGAVVVGAVMAAAMPFSYGIIRSSVLILOVPEPLVTVSGADNIFIVLE	720
Db	661	MVDSKATLGLGGAVVVGAVMAAAMPFSYGIIRSSVLILOVPEPLVTVSGADNIFIVLE	720
QY	721	YQRLPRRPGEBREYHIGRALGRVAPSMULCSLSFAICFTFGALTMPAVRTFALTSGLAV	780
Db	721	YQRLPRRPGEBREYHIGRALGRVAPSMULCSLSFAICFTFGALTMPAVRTFALTSGLAV	780
QY	781	ILDLLOMSAFVALSLDCKROBARLDVCCCVCPQGLPRPGQEGSLLGPOKAVAPFL	840
Db	781	ILDLLOMSAFVALSLDCKROBARLDVCCCVCPQGLPRPGQEGSLLGPOKAVAPFL	840
QY	841	LHMHTTRGVVLLFLALFGVSLYSMCHISVGLDQELAPKOSYLLDYLFLNRYEYEVAPV	900
Db	841	LHMHTTRGVVLLFLALFGVSLYSMCHISVGLDQELAPKOSYLLDYLFLNRYEYEVAPV	900
QY	901	YFVTTLLGYNPSSSEAGNNAICSSAGCNPNSFTQKIQVATBPEQSYLAIPASSWYDFTDW	960
Db	901	YFVTTLLGYNPSSSEAGNNAICSSAGCNPNSFTQKIQVATBPEQSYLAIPASSWYDFTDW	960
QY	961	LTBSSCCRLYISGNNKQKFCPSTYNSLNCIKNCSITMGSVRPBEQFHYLLPWFLLDRP	1020
Db	961	LTBSSCCRLYISGNNKQKFCPSTYNSLNCIKNCSITMGSVRPBEQFHYLLPWFLLDRP	1020
QY	1021	NIKCPKGGALAYSPSNVLTSQGVLASRPMAYHKLKNSODYTEALPAARELANITRADL	1080
Db	1021	NIKCPKGGALAYSPSNVLTSQGVLASRPMAYHKLKNSODYTEALPAARELANITRADL	1080
QY	1081	RKVGDTDPAPFVFPYTTITNVFBEQYLTILPEGLFMLSCLVPTFAVSCLLGLDLRSGLL	1140
Db	1081	RKVGDTDPAPFVFPYTTITNVFBEQYLTILPEGLFMLSCLVPTFAVSCLLGLDLRSGLL	1140
QY	1141	NLSISVIMLVDTVGFMAIMDISYNAVSLINLVSANVGSVEFVSHITTSFAISTPYMLER	1200
Db	1141	NLSISVIMLVDTVGFMAIMDISYNAVSLINLVSANVGSVEFVSHITTSFAISTPYMLER	1200
QY	1201	AKEXTTMSGAVFVGAVMTNLPGLIVGLAKAOLIOIFPRILNLTLLTGLLHGVLFPV	1260
Db	1201	AKEXTTMSGAVFVGAVMTNLPGLIVGLAKAOLIOIFPRILNLTLLTGLLHGVLFPV	1260
QY	1261	ILSYVGDDVNPALALEQKRAEEAAVAAIMVASCPNHPBRVGTADNIYVNHSEEGSIKGA	1320

Db 1261 ILSYVDPVNPALALQKRAEEAVAAVMVASCPNHSRVSATDNIVYNSHSPBSIGKAGAA 1320
 QY 1321 ISNFLPNNGROF 1332
 1321 ISNFLPNNGROF 1332

RESULT 6
 AAM79168
 ID AAM79168 standard; protein; 1359 AA.

AC AAM79168;

DT 06-NOV-2001 (first entry)

DE Human protein SEQ ID NO 1830.

KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation.

OS Homo sapiens.

PN WO200157190-A2.

PD 09-AUG-2001.

PF 05-FEB-2001; 2001MO-US004098.

PR 03-FEB-2000; 2000US-00496914.

PR 27-APR-2000; 2000US-00560875.

PR 20-JUN-2000; 2000US-00598075.

PR 19-JUL-2000; 2000US-00620325.

PR 01-SEP-2000; 2000US-00654936.

PR 15-SEP-2000; 2000US-00663561.

PR 20-OCT-2000; 2000US-00693325.

PR 30-NOV-2000; 2000US-00728422.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y,
 PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
 PI Xue AJ, Yang Y, Wejrtman T, Goodrich R;

DR WPI; 2001-476283/51.

DR N-PSDB; AAK52301.

PT Nucleic acids encoding polypeptides with cytokine-like activities, useful
 PT in diagnosis and gene therapy.

PS Claim 20; Page 4213-4216; 6221pp; English.

CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAM78333-AAK80302) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activity/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
 CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
 CC sequence listing were missing at the time of publication

XX Sequence 1359 AA;

Query Match 99.5%; Score 6872.5; DB 4; Length 1359;
 Best Local Similarity 97.9%; Pred. NO. 0;
 Matches 1330; Conservative 0; Mismatches 2; Indels 27; Gaps 1;

QY 1 MAEAGLRGWLMLALLRLAQSEPYTTIHQPGYCAFYDECGKNBELSGSLMTLSNVSCLSN 60
 Db 1 MAEAGLRGWLMLALLRLAQSEPYTTIHQPGYCAFYDECGKNBELSGSLMTLSNVSCLSN 60
 QY 61 TPAKRTTGHLLILQKICRPLTYGPTQACCSKQVLSLEASITKALLTRCPACSDNF 120
 Db 61 TPAKRTTGHLLILQKICRPLTYGPTQACCSKQVLSLEASITKALLTRCPACSDNF 120
 QY 121 VNLHCNHTCSPNOSLINTVTRVAQLGAGQLPAVVAEAFYQSHFPAQSYDCSRVVPAA 180
 Db 121 VNLHCNHTCSPNOSLINTVTRVAQLGAGQLPAVVAEAFYQSHFPAQSYDCSRVVPAA 180
 QY 181 ATLAAGTCGVVGSALCNARWLNFGQDTGNGIAPLDITPHLLPQAVSGSIOPLNEGV 240
 Db 181 ATLAAGTCGVVGSALCNARWLNFGQDTGNGIAPLDITPHLLPQAVSGSIOPLNEGV 240
 QY 241 ARCNEOGDDVATCSQDCAASCPAIPARQALDSTYVLGQMPQSLVLIILCSVFAVNTI 300
 Db 241 ARCNEOGDDVATCSQDCAASCPAIPARQALDSTYVLGQMPQSLVLIILCSVFAVNTI 300
 QY 301 LTVGFVAPARADSKMVDPRKGTSLSDKLSFSTHTLLGQFPQGMGTWVASMPLTIVLSV 360
 Db 301 LTVGFVAPARADSKMVDPRKGTSLSDKLSFSTHTLLGQFPQGMGTWVASMPLTIVLSV 360
 QY 361 IPVVALAAGLVPTLTTDEVELMSAPNSQARSEKAFHDQFGPFRTNQVILTAPNRSY 420
 Db 361 IPVVALAAGLVPTLTTDEVELMSAPNSQARSEKAFHDQFGPFRTNQVILTAPNRSY 420
 QY 421 RYDLSLLGPRNFGSIIIDDLLELELQERLRLQVSWPBAQRNLSQDICTAPLPNDT 480
 Db 421 RYDLSLLGPRNFGSIIIDDLLELELQERLRLQVSWPBAQRNLSQDICTAPLPNDT 480
 QY 481 SLVDCINSILQYFQNNRTLLLTANQTLMGQTSQVDMQDHPFYCANAPLTFQDGTALAL 540
 Db 481 SLVDCINSILQYFQNNRTLLLTANQTLMGQTSQVDMQDHPFYCANAPLTFQDGTALAL 540
 QY 541 SCWADYGAVFPFLAIGYKGYKGYSEAEALIMTFSLNTPAGDPRLAQAALMEAEFLBEM 600
 Db 541 SCWADYGAVFPFLAIGYKGYKGYSEAEALIMTFSLNTPAGDPRLAQAALMEAEFLBEM 600
 QY 601 RAFQRMAGMFOVFTTAERSLEDEINRTTAEDLPITATSYIVFLYISLALSGSYSSKRV 660
 Db 601 RAFQRMAGMFOVFTTAERSLEDEINRTTAEDLPITATSYIVFLYISLALSGSYSSKRV 660
 QY 661 MVDSKATLIGAGVAVLVGAVMAMGFFSTUGIRSSLVIIQVVEFLVLSGADNIFIFVLE 720
 Db 661 MVDSKATLIGAGVAVLVGAVMAMGFFSTUGIRSSLVIIQVVEFLVLSGADNIFIFVLE 720
 QY 721 YQRLPRRPGEPREVIHGRALGRVAPSMILCSISEAICFPLGALTMPAVRTFALTGLAV 780
 Db 721 YQRLPRRPGEPREVIHGRALGRVAPSMILCSISEAICFPLGALTMPAVRTFALTGLAV 780
 QY 781 IIDFLLQMSAFVALSLDSKROEASRLDVCCCYKPOELPPPGGGBGLLGFQKAYAPFL 840
 Db 781 IIDFLLQMSAFVALSLDSKROEASRLDVCCCYKPOELPPPGGGBGLLGFQKAYAPFL 840
 QY 841 LHWITGVVLLFLALFGVSLVSMCHISVGLQDELALPDOSYLDVFLINRFEVGA 900
 Db 841 LHWITGVVLLFLALFGVSLVSMCHISVGLQDELALPDOSYLDVFLINRFEVGA 900
 QY 901 YFVTTIGVYFSSBAGNNAICSSAGCNFFSTQKIQTATFPPEQSYLAIPASSVVDPFIDW 960
 Db 901 YFVTTIGVYFSSBAGNNAICSSAGCNFFSTQKIQTATFPPEQSYLAIPASSVVDPFIDW 960
 QY 961 LTPSSCCRLYISGPNKDCPSTVNSLNCIKNCMSITMGSVBPVSQFHXYLFWPFLNDRP 1020
 Db 961 LTPSSCCRLYISGPNKDCPSTVNSLNCIKNCMSITMGSVBPVSQFHXYLFWPFLNDRP 1020
 QY 1021 NIKCPKGGIAAYSTVNLTSQGYL-----ASRPMAYH 1053
 Db 1021 NIKCPKGGIAAYSTVNLTSQGYLDTVAILSPRLYSGTISAHCNLYLLDSASRPMAYH 1080

QY 1054 KPLKKSQDYTEALRAARELAANTADLRKVPCTDPAFEVFPYTTINVFPEOYLTIPEGL 1113
 DB 1081 KPLKKSQDYTEALRAARELAANTADLRKVPCTDPAFEVFPYTTINVFPEOYLTIPEGL 1140
 QY 1114 FMLSICLVTFPAVSCILGLDLSRGLNLTSMILVDVVGFMALMDISYNAVSLINLVS 1173
 DB 1141 FMLSICLVTFPAVSCILGLDLSRGLNLTSMILVDVVGFMALMDISYNAVSLINLVS 1200
 QY 1174 AVGMSVEFVSHITRSPFAISTKPTWLERAKEBATISMSAVFAGVAMTNLPGLVLGLAKAQ 1233
 DB 1201 AVGMSVEFVSHITRSPFAISTKPTWLERAKEBATISMSAVFAGVAMTNLPGLVLGLAKAQ 1260
 QY 1234 LLIQIFPFRNLNLTLLGLHGLVFLPYIISYGPDVNPLALBQKAEBAVAAVMAVASC 1293
 DB 1261 LLIQIFPFRNLNLTLLGLHGLVFLPYIISYGPDVNPLALBQKAEBAVAAVMAVASC 1320
 QY 1294 NHPSRVSTADNTYVNHSPFSGISIKGAGAINFPLPNNGROF 1332
 DB 1321 NHPSRVSTADNTYVNHSPFSGISIKGAGAINFPLPNNGROF 1359
 RESULT 7
 ADJ27337
 ID ADJ27337 standard; protein; 1359 AA.
 AC ADJ27337;
 DT 20-MAY-2004 (first entry)
 XX
 DE Human NPC1L1.
 KW Niemann-Pick disease; type C1; gene-like 1; NPC1L1; trans-golgi network;
 KW plasma membrane; transport signal; promoter;
 KW sterol regulated element binding protein 1; SREBP1;
 KW binding consensus sequence; transmembrane domain; sterol-sensing domain;
 KW SSD; cholesterol; NPC1; receptor; Niemann-Pick C1 disease; intestinal;
 KW cholesterol absorption; serum cholesterol; hyperlipidaemia;
 KW atherosclerosis; coronary heart disease; stroke; arteriosclerosis.
 XX
 OS Homo sapiens.
 PN WO2004009772-A2.
 XX
 PD 29-JAN-2004.
 XX
 PF 17-JUL-2003; 2003WO-US022467.
 XX
 PR 19-JUL-2002; 2002US-0397442P.
 XX
 PA (SCHER) SCHERING CORP.
 PI Altmann SW, Murgolo NJ, Wang LQ, Graziano MP;
 XX
 DR MPI: 2004-132945/13.
 DR N-PSDB; ADJ27336.
 DR GENBANK; AF192522.
 XX
 PT New Niemann-Pick disease, type C1, gene-like 1 polypeptide, for detecting
 PT or identifying antagonists of NPC1L1 for inhibiting intestinal
 PT cholesterol absorption in a subject, or for treating elevated serum
 PT cholesterol or stroke.
 XX
 PS Example 7; SEQ ID NO 44; 125bp; English.
 XX
 CC This sequence represents a Niemann-Pick disease, type C1, gene-like 1
 CC (NPC1L1) polypeptide. NPC1L1 is an N-glycosylated protein which contains
 CC a motif ADJ27331 which acts as a trans-golgi network to plasma membrane
 CC transport signal, and which exhibits limited tissue distribution and
 CC gastrointestinal abundance. The human NPC1L1 promoter sequence contains a
 CC sterol regulated element binding protein 1 (SREBP1) binding consensus
 CC sequence. NPC1L1 has 13 transmembrane spanning segments and a sterol-
 CC sensing domain (SSD) which is involved in sensing cholesterol levels,
 CC possibly by a mechanism which involves direct cholesterol bonding. NPC1L1

CC has 42% amino acid sequence homology to human NPC1, a receptor
 CC responsible for Niemann-Pick C1 disease. The NPC1L1 polypeptides or
 CC polynucleotides are useful for detecting or identifying antagonists of
 CC NPC1L1, which can be used for inhibiting intestinal cholesterol
 CC absorption in a subject, or for treating medical conditions including
 CC elevated serum cholesterol, hyperlipidaemia, atherosclerosis, coronary
 CC heart disease, stroke or arteriosclerosis.
 XX
 SQ Sequence 1359 AA;
 Query Match 99.5%; Score 6872.5; DB 8; Length 1359;
 Best Local Similarity 97.9%; Pred. No. 0;
 Matches 1330; Conservative 0; Mismatches 2; Indels 27; Gaps 1;
 QY 1 MAEAGIRGMLWALLRLAQSSEPYTTIHQPGYCAFYDCGKXNPLSGSLMTLSNVSCLSN 60
 DB 1 MAEAGIRGMLWALLRLAQSSEPYTTIHQPGYCAFYDCGKXNPLSGSLMTLSNVSCLSN 60
 QY 61 TPARKITGDHLLLOKICPRLYTGPNTQACCAKQVLSLEASLSTTKALLTRCPACSDNF 120
 DB 61 TPARKITGDHLLLOKICPRLYTGPNTQACCAKQVLSLEASLSTTKALLTRCPACSDNF 120
 QY 121 VNLHCNHTCSPNQSLFINTVTRVAQAGAGOLPAVVAYEAFYQHSFAEOSYDSCSRVRVPA 180
 DB 121 VNLHCNHTCSPNQSLFINTVTRVAQAGAGOLPAVVAYEAFYQHSFAEOSYDSCSRVRVPA 180
 QY 181 ATLAVGTMGVYGSAALCNAQRWLNFGDPTGNGLABLDITFHLLEBQAVAGSIIQPLNEGV 240
 DB 181 ATLAVGTMGVYGSAALCNAQRWLNFGDPTGNGLABLDITFHLLEBQAVAGSIIQPLNEGV 240
 QY 241 AHCNSQGDVATCGCCDCAASCRAIARPOLDSTFFYIGQMGSVLITILCSVAVVTI 300
 DB 241 AHCNSQGDVATCGCCDCAASCRAIARPOLDSTFFYIGQMGSVLITILCSVAVVTI 300
 QY 301 LTVGRVAPARPKSKMVPKKGTSLSDKLSEFTHTLLGQFQSGMTWVASWPLTILVLSV 360
 DB 301 LTVGRVAPARPKSKMVPKKGTSLSDKLSEFTHTLLGQFQSGMTWVASWPLTILVLSV 360
 QY 361 IPVVALAAGLVFTLTDTPVELMSAPNSQARSEKAFHDQHPFPFRITNVITAPNRSY 420
 DB 361 IPVVALAAGLVFTLTDTPVELMSAPNSQARSEKAFHDQHPFPFRITNVITAPNRSY 420
 QY 421 RYDSLLGPKNFGSLIDDLLELLELOERLRLQVSPBEAQRNISTODICTAPLNPNPT 480
 DB 421 RYDSLLGPKNFGSLIDDLLELLELOERLRLQVSPBEAQRNISTODICTAPLNPNPT 480
 QY 481 SYDCCINSILOYPQNNRTLLLTANQTLMGQTSQVDMKDHFLYCANNAPLTRKDTALAL 540
 DB 481 SYDCCINSILOYPQNNRTLLLTANQTLMGQTSQVDMKDHFLYCANNAPLTRKDTALAL 540
 QY 541 SCMDYGAAPVPPFLAIGYKGDYSEAEALINTFSLNNYPAGDPRLAQAKLMEBAFLEBM 600
 DB 541 SCMDYGAAPVPPFLAIGYKGDYSEAEALINTFSLNNYPAGDPRLAQAKLMEBAFLEBM 600
 QY 601 RAFORRMAGMPQVTTAERSLDEINRTTAEBLPIFATSYIYIFLYISLALGSYSWSRV 660
 DB 601 RAFORRMAGMPQVTTAERSLDEINRTTAEBLPIFATSYIYIFLYISLALGSYSWSRV 660
 QY 661 WYDSKATLGLGVAVALGAVMAAMGFSGYLSGRSSLVILQVVPFLVLSVAGADNIFIVYLE 720
 DB 661 WYDSKATLGLGVAVALGAVMAAMGFSGYLSGRSSLVILQVVPFLVLSVAGADNIFIVYLE 720
 QY 721 YORLRPRRPEPRVHIGRALGRVABSMILCSISEAICFFLGALTMPAVRTFALTSGLA 780
 DB 721 YORLRPRRPEPRVHIGRALGRVABSMILCSISEAICFFLGALTMPAVRTFALTSGLA 780
 QY 781 IUDFLQMSAFALLSLDSKROEASRLDYCCVYKQBELPPRQGGGLLIGFQKXAPPL 840
 DB 781 IUDFLQMSAFALLSLDSKROEASRLDYCCVYKQBELPPRQGGGLLIGFQKXAPPL 840
 QY 841 LHMTRGVTLILFLALFGVSLYSKCHISVGLDQELALPDSDYLLYPLFLNRYFEVGA 900
 DB 841 LHMTRGVTLILFLALFGVSLYSKCHISVGLDQELALPDSDYLLYPLFLNRYFEVGA 900

QY 901 YFVTTIGVNFSSBAGNNAICSSAGCNNSFTQKIQVATEPPEQSYAIAPASSWVDDFTDM 960
DB 901 YFVTTIGVNFSSBAGNNAICSSAGCNNSFTQKIQVATEPPEQSYAIAPASSWVDDFTDM 960
QY 961 LPPSSCCRLYISGPNKDKCPSTVNSLNCNKCMSITMGSVRSVQFHKYLPWFINDRP 1020
DB 961 LPPSSCCRLYISGPNKDKCPSTVNSLNCNKCMSITMGSVRSVQFHKYLPWFINDRP 1020
QY 1021 NIKCPKGLAAVSTVNLTSDDQVL-----ASRPMAYH 1053
DB 1021 NIKCPKGLAAVSTVNLTSDDQVL-----ASRPMAYH 1053
QY 1054 KPLKNSQDYTEALRAARELANITADLRKVPDTPAEFVPPYTTINVFYEQYITLPEGI 1113
DB 1081 KPLKNSQDYTEALRAARELANITADLRKVPDTPAEFVPPYTTINVFYEQYITLPEGI 1140
QY 1114 FMLSCLVPTPAVSCILLGLDLRSGLNLSTMIIVDVVGFPMALDISYNAVSLINYS 1173
DB 1141 FMLSCLVPTPAVSCILLGLDLRSGLNLSTMIIVDVVGFPMALDISYNAVSLINYS 1200
QY 1174 AVGMSVEFVSHITRSPAISTKPTWLERAKEATISMGSAVFAVAMTNLPGLVGLAKAQ 1233
DB 1201 AVGMSVEFVSHITRSPAISTKPTWLERAKEATISMGSAVFAVAMTNLPGLVGLAKAQ 1260
QY 1234 LQIPEFRNLITLLGLHGLVFLPVIIISYVGPDVNPALALEQKAEAAVAAVWVASCP 1293
DB 1261 LQIPEFRNLITLLGLHGLVFLPVIIISYVGPDVNPALALEQKAEAAVAAVWVASCP 1320
QY 1294 NHRSRVSTADNTYVNHSPFGSITKAGAGISNPLPNNRQF 1332
DB 1321 NHRSRVSTADNTYVNHSPFGSITKAGAGISNPLPNNRQF 1359
RESULT 8
ID ADY60780 standard; protein, 1359 AA.
AC ADY60780;
XX
DT 19-MAY-2005 (first entry)
XX
DE Human Niemann-Pick C1-like protein 1 (NPC1L1) protein SEQ ID 44.
XX
NIemann-Pick C1-like protein 1; Antilipemic; Antiartherosclerotic;
XX
KM Cardiac; Vasoconstrictive; Cerebroprotective; cholesterol; hyperlipidemia;
XX
KM metabolic disorder; atherosclerosis; cardiovascular disease;
XX
KM coronary artery disease; cerebrovascular ischemia; neurological disease;
XX
KM arteriosclerosis; transgenic animal.
OS Homo sapiens.
XX
PN WO2005015968-A1.
XX
PD 24-FEB-2005.
XX
PF 16-DEC-2003; 2003WO-US040113.
XX
PR 17-JUL-2003; 2003US-00621758.
XX
PR 22-AUG-2003; 2003US-00646301.
XX
PR 16-SEP-2003; 2003US-00663208.
XX
PA (SCHI) SCHERING CORP.
XX
PI Altmann SM, Murgolo NJ, Wang LQ, Graziano MP;
XX
DR WPI; 2005-284403/29.
XX
DR N-PSDB; ADY60779.
PT Novel isolated NPC1L1 polypeptide, useful for identifying antagonist of
PT NPC1L1 that inhibits NPC1L1 mediated sterol or S alpha-sterol uptake, for
XX treating hyperlipidemia or atherosclerosis in mouse, rat or human.
XX

PS Disclosure; SEQ ID NO 44; 146bp; English.
XX
XX The invention relates to an isolated Niemann-Pick C1-like protein 1
CC (NPC1L1) polypeptide comprising 42 or more contiguous amino acids from an
CC amino acid sequence appearing as ADY60734, ADY60740 or ADY60748 (Rat,
CC human and Mouse NPC1L1). Also included are an isolated polynucleotide
CC encoding NPC1L1 (appearing as ADY60737, ADY60739 or ADY60747), a
CC recombinant vector comprising the polynucleotide, a host cell comprising
CC the vector, an isolated antibody which specifically binds to NPC1L1 (or
CC to a rat NPC1L1 peptide sequence appearing as ADY60775-ADY60778),
CC preparation of NPC1L1, a mutant transgenic mouse comprising a homozygous
CC mutation of endogenous chromosomal NPC1L1 (where the mouse does not
CC produce any functional NPC1L1 protein), an offspring or progeny of the
CC mouse having inherited a mutated NPC1L1 allele of the mouse, screening a
CC sample for an intestinal sterol or Salpa-sterol absorption antagonist,
CC inhibiting NPC1L1 mediated sterol or Salpa-sterol uptake in a subject, a
CC kit (comprising ezetimibe or substituted ezetimibe in a pharmaceutical
CC dosage form, and information indicating that NPC1L1 is a target of
CC ezetimibe or substituted ezetimibe), decreasing the level of intestinal
CC sterol or Salpa-sterol absorption in a subject (involving reducing the
CC level of expression of NPC1L1 in the subject), identifying an antagonist
CC of NPC1L1 and an isolated mammalian cell which lacks a gene encoding a
CC functional NPC1L1 protein. NPC1L1 is useful for identifying an antagonist
CC of NPC1L1 which is useful for inhibiting or decreasing the level of
CC NPC1L1 mediated sterol or Salpa-sterol uptake in a subject such as
CC mouse, rat or human. The level of expression of NPC1L1 in the subject is
CC reduced by mutating NPC1L1 in the subject. The NPC1L1 knockout mouse is
CC useful for screening a sample for intestinal sterol or Salpa-sterol
CC absorption antagonist. The NPC1L1 antagonists identified are useful for
CC treating medical conditions caused or mediated by NPC1L1, e.g.,
CC hyperlipidemia, atherosclerosis, coronary heart diseases, stroke or
CC arteriosclerosis. The present sequence represents a mammalian NPC1L1
XX protein.
XX
SQ Sequence 1359 AA:
Query Match 99.5%; Score 6872.5; DB 9; Length 1359;
Best Local Similarity 97.9%; Pred. No. 0;
Matches 1330; Conservative 0; Mismatches 2; Indels 27; Gaps 1;
QY 1 MABAGRGWLMWALRLAOSRPYTHQPGCAFYDECKRNPRLSGSLMTSNVSCSN 60
DB 1 MABAGRGWLMWALRLAOSRPYTHQPGCAFYDECKRNPRLSGSLMTSNVSCSN 60
QY 61 TPARKITGPHLLIKICPRLTYGPNTOACSAKOLVSHASLITKALLTRCPACSDNF 120
DB 61 TPARKITGPHLLIKICPRLTYGPNTOACSAKOLVSHASLITKALLTRCPACSDNF 120
QY 121 VNIHCHNTSPNQSIFINVTVAQLGAGQLPAVVAVEAFYQHSFPAQSYDSCSRVPPAA 180
DB 121 VNIHCHNTSPNQSIFINVTVAQLGAGQLPAVVAVEAFYQHSFPAQSYDSCSRVPPAA 180
QY 181 ATLAVGTMGVYGSALCNAQRWLNFOGDTGNGLAFLDITPHLLBERGQAVGSGIOPLANEBV 240
DB 181 ATLAVGTMGVYGSALCNAQRWLNFOGDTGNGLAFLDITPHLLBERGQAVGSGIOPLANEBV 240
QY 241 ARCNSQGDVATCGCODCAACPAIARPOALDSTFYIGQMGSLVLIITLCSVPAVVTI 300
DB 241 ARCNSQGDVATCGCODCAACPAIARPOALDSTFYIGQMGSLVLIITLCSVPAVVTI 300
QY 301 LLVGFVAVAPARDKSKVNDPKKGTSLDKLSFSTHTLLGQFQGWGTWASWPLTILVLSV 360
DB 301 LLVGFVAVAPARDKSKVNDPKKGTSLDKLSFSTHTLLGQFQGWGTWASWPLTILVLSV 360
QY 361 IPVVALAAGLVTELTTPVEIWSAPNSQABEKAFOHQHBPFRPTNOVITLTPRRSY 420
DB 361 IPVVALAAGLVTELTTPVEIWSAPNSQABEKAFOHQHBPFRPTNOVITLTPRRSY 420
QY 421 RYDSLLGPKNFSGLIDLLLELELEORLHLQVWSEAOGRNISLQDICYAPLNPDT 480
DB 421 RYDSLLGPKNFSGLIDLLLELELEORLHLQVWSEAOGRNISLQDICYAPLNPDT 480
QY 481 SLVDCINSLLOYPONRTLLLTANQTLMGQTSQVDMKDHFLYCANNAPLTKDGTALAL 540


```

Db      481  |||...||| 540
      481  SLYDCCINSLLQYFNNRTLLLTANQTMGQTSQVMDXHLFCANAPLTKDGTALL
Qy      541  SCMDYGA PVFPPLAIGYKGYSEAEALIMTFSLNNYPADPRLAOKLMEAEFLBEM 600
      541  SCMDYGA PVFPPLAIGYKGYSEAEALIMTFSLNNYPADPRLAOKLMEAEFLBEM 600
Qy      601  RAFOORMAGMFOVTTAERSLEDEINRTTAEDLPFATSIVIFLYISLALGSYSMSRV 660
      601  RAFOORMAGMFOVTTAERSLEDEINRTTAEDLPFATSIVIFLYISLALGSYSMSRV 660
Qy      661  MVDKATLIGGVAVVILGAVMAAMGFPSYLGIRSSVLIVQVVPFLVLSGANIFPVLE 720
      661  MVDKATLIGGVAVVILGAVMAAMGFPSYLGIRSSVLIVQVVPFLVLSGANIFPVLE 720
Qy      721  YQRLRRRGEPRREVIHGRALGRVAPSMILCSLSEALCFPLGALTMPAVRTALTSGLAV 780
      721  YQRLRRRGEPRREVIHGRALGRVAPSMILCSLSEALCFPLGALTMPAVRTALTSGLAV 780
Qy      781  ILDFLLQMSAFVALLSLDSKQGBASRLDYCCCVKQPELPPQGGEGLLGFPQKAVAPFL 840
      781  ILDFLLQMSAFVALLSLDSKQGBASRLDYCCCVKQPELPPQGGEGLLGFPQKAVAPFL 840
Qy      841  LHMTRGVVLLFLALFGVSLYSMCHISVGLDQELAPKDSYLLDYFLFLNRYFVGAVY 900
      841  LHMTRGVVLLFLALFGVSLYSMCHISVGLDQELAPKDSYLLDYFLFLNRYFVGAVY 900
Qy      901  YVVTLLGVNFSSEAGNNAICSGAGCNFFSFTOKIQATFPPQSYLAIPASSMVDFFDM 960
      901  YVVTLLGVNFSSEAGNNAICSGAGCNFFSFTOKIQATFPPQSYLAIPASSMVDFFDM 960
Qy      961  LTPSSCCRLYISGPNKDFCPSSTVNSLNCNKMSITMGSVRPSVEQPHKYLPMFLNDP 1020
      961  LTPSSCCRLYISGPNKDFCPSSTVNSLNCNKMSITMGSVRPSVEQPHKYLPMFLNDP 1020
Qy      1021  NIKCKRGGLAASTSVNLTSQVLT-----ASBEMAH 1053
      1021  NIKCKRGGLAASTSVNLTSQVLT-----ASBEMAH 1053
Qy      1054  KPLKNSQDYTEALRAARELANITADLRKVPGTDPAFEVFPYITINVFEGYLLTLEGL 1113
      1054  KPLKNSQDYTEALRAARELANITADLRKVPGTDPAFEVFPYITINVFEGYLLTLEGL 1113
Qy      1114  FMLSICLVPTFAVSCLLGLDLRSGLNLISIMLIVDTVGFMALMGSINAVSLINLVS 1173
      1141  FMLSICLVPTFAVSCLLGLDLRSGLNLISIMLIVDTVGFMALMGSINAVSLINLVS 1200
Qy      1174  AVGMSEVEFVSHITRFAISTKPTWLERAKEATISMGSAVFAGVAMTNLPGIILVGLAKQ 1233
      1201  AVGMSEVEFVSHITRFAISTKPTWLERAKEATISMGSAVFAGVAMTNLPGIILVGLAKQ 1260
Qy      1234  LIQIFFFRNLITLLGLHGLVFLPVILSYGPDVNPALALBOKRAEAVAAVWVASC 1293
      1261  LIQIFFFRNLITLLGLHGLVFLPVILSYGPDVNPALALBOKRAEAVAAVWVASC 1320
Qy      1294  NHPSEVSTADNITYNHSFEGSIKGAGAISNPLPNNGRQF 1332
      1321  NHPSEVSTADNITYNHSFEGSIKGAGAISNPLPNNGRQF 1359
Db

```

```

XX      OS      Homo sapiens.
XX      PN      WO2005069900-A2.
XX      PD      04-AUG-2005.
XX      PF      14-JAN-2005; 2005WO-US001469.
XX      PR      16-JAN-2004; 2004US-0537341P.
XX      PA      (MERI ) MERCK & CO INC.
XX      PI      Garcia-Calvo M;
XX      DR      WPI; 2005-564070/57.
XX      DR      N-PSDB; AEB93610.
XX      DR      GENBANK; AF192522.
XX      PT      Identifying ligand of NPC1L1 for stimulating the activity of NPC1L1, by
XX      PT      contacting NPC1L1 with detectably labeled substituted 2-azetidione
XX      PT      glucuronide and a candidate compound and determining if compound binds to
XX      PT      human NPC1L1.
XX      PS      Disclosure; SEQ ID NO 44; 215pp; English.
XX      CC      The invention relates to identifying a ligand of NPC1L1. The method
XX      CC      involves contacting human NPC1L1 with a detectably labeled substituted 2-
XX      CC      azetidione glucuronide and a candidate compound and determining whether
XX      CC      the candidate compound binds to human NPC1L1. In identifying a ligand of
XX      CC      NPC1L1, the KD value is 100 nM or lower, preferably 50 nM, 20 nM, or 10
XX      CC      nM or lower. The detectably labeled substituted 2-azetidione glucuronide
XX      CC      is 35 S-labeled compound 2, given in the specification. NPC1L1 ligands
XX      CC      are useful for stimulating or blocking the activity of NPC1L1, and for
XX      CC      treating conditions caused or mediated by NPC1L1. It is useful for
XX      CC      reducing the incidence of hyperlipidemia, atherosclerosis, coronary heart
XX      CC      disease, stroke, or arteriosclerosis. The present sequence represents the
XX      CC      amino acid sequence of human NPC1L1 as disclosed under Genbank accession
XX      CC      number AF192522.
XX      SQ      Sequence 1359 AA:
XX
Query Match      99.5%; Score 6872.5; DB 9; Length 1359;
Beet Local Similarity 97.9%; Pred. No. 0;
Matches 1330; Conservative 0; Mismatches 2; Indels 27; Gaps 1;
Qy      1  MAEAGLRGMLMALIRLAQSEPYTHIQPGCAFYDECGNPKLSGLMTLSNVCSLSN 60
      1  MAEAGLRGMLMALIRLAQSEPYTHIQPGCAFYDECGNPKLSGLMTLSNVCSLSN 60
Db      61  TPARKITGDHILLOKICPRLYTGPNTOACCSAKQLVSLBASLSTTKALLTRCPACSDNF 120
      61  TPARKITGDHILLOKICPRLYTGPNTOACCSAKQLVSLBASLSTTKALLTRCPACSDNF 120
Qy      121  VNLHCNHTCSRNQSLFIVNTRVAQLGAGQLPAVVAAYEAFYQHSFPAEBSYDSCSRVRVAA 180
      121  VNLHCNHTCSRNQSLFIVNTRVAQLGAGQLPAVVAAYEAFYQHSFPAEBSYDSCSRVRVAA 180
Qy      181  ATLAVTMCGVYGSAALCAQRMLNFGDGTNGLAFLDITFHLBERGQAVGSGIOPLNCGV 240
      181  ATLAVTMCGVYGSAALCAQRMLNFGDGTNGLAFLDITFHLBERGQAVGSGIOPLNCGV 240
Qy      241  ARCNESQGDVAVTSCQDCAASCPIARPAQLDSTFYLGONPGSLVLIILICSVAVVTI 300
      241  ARCNESQGDVAVTSCQDCAASCPIARPAQLDSTFYLGONPGSLVLIILICSVAVVTI 300
Db      301  LTVGRVAPARDKSMVDPKKGTSLSDKLSSTHTLLGQFQCGWGTWASWPLTILVLSV 360
      301  LTVGRVAPARDKSMVDPKKGTSLSDKLSSTHTLLGQFQCGWGTWASWPLTILVLSV 360
Qy      361  IPVVALAAGLVFTLTTTPVELMSAPNSQASSEKAFHDOHGPFRTQOVILTANRSSY 420
      361  IPVVALAAGLVFTLTTTPVELMSAPNSQASSEKAFHDOHGPFRTQOVILTANRSSY 420
Db

```

RESULT 9
 AEB93610
 ID AEB93610 standard; protein; 1359 AA.
 XX
 AC AEB93610;
 XX
 DT 06-OCT-2005 (first entry)
 XX
 DE Human NPC1L1 protein.
 XX
 KW Protein engineering; NPC1L1; antidiabetic; antiarteriosclerotic;
 KW carotid; vasotropic; hyperlipidemia; atherosclerosis; stroke;
 KW arteriosclerosis; ds.


```

QY 421 RYDSLLGPKNFSGLLDLLELLELQERLRHQAWSBEAQNISLODICVAPLNPDMT 480
DB 421 RYDSLLGPKNFSGLLDLLELLELQERLRHQAWSBEAQNISLODICVAPLNPDMT 480
QY 481 SLYDCCINSLLQYFQNNRLLLTITANQTMGQTSQVDMKDHFLYCANAPLTFKDGATALL 540
DB 481 SLYDCCINSLLQYFQNNRLLLTITANQTMGQTSQVDMKDHFLYCANAPLTFKDGATALL 540
QY 541 SCMAADGAVPEPFLAIGYKKGQYSEABALIMFESLNNTYPAQBPRLAQAALMEAELEEM 600
DB 541 SCMAADGAVPEPFLAIGYKKGQYSEABALIMFESLNNTYPAQBPRLAQAALMEAELEEM 600
QY 601 RAFORAMAGMFOYTFPABRSLEDEINRTTAAEDLPFATSYIVFLYISLALGSSYSMSRY 660
DB 601 RAFORAMAGMFOYTFPABRSLEDEINRTTAAEDLPFATSYIVFLYISLALGSSYSMSRY 660
QY 661 MDSKATLIGLGVAVVLGAVMAAMGPFSSYLGRSSLVILQVDFVLVSQADNIFIFVLE 720
DB 661 MDSKATLIGLGVAVVLGAVMAAMGPFSSYLGRSSLVILQVDFVLVSQADNIFIFVLE 720
QY 721 YORLPRRPGEPREPVHIGRALGRAPSMILCSSEALCFPLGALTMPAVRTFALTSGLAV 780
DB 721 YORLPRRPGEPREPVHIGRALGRAPSMILCSSEALCFPLGALTMPAVRTFALTSGLAV 780
QY 781 ILDFLLQMSAFVALLSLDSKROBASRLDYCCCVKPOELPPPGQEGELLGFFQKAYAPFL 840
DB 781 ILDFLLQMSAFVALLSLDSKROBASRLDYCCCVKPOELPPPGQEGELLGFFQKAYAPFL 840
QY 841 LHMITRGVALLFLFALFGVLSYMGCHISVGLDQELALPKDSYLLDYFLPLNRYFEVGA 900
DB 841 LHMITRGVALLFLFALFGVLSYMGCHISVGLDQELALPKDSYLLDYFLPLNRYFEVGA 900
QY 901 YPFTTIGVNPSSSAGMNAICSSAGCNPSFTQKIQVATEPEBSYLAIPASSVNDPEIDM 960
DB 901 YPFTTIGVNPSSSAGMNAICSSAGCNPSFTQKIQVATEPEBSYLAIPASSVNDPEIDM 960
QY 961 LPPSSCCRLYISGPNKDKECPSTVNSLNCJKNCSITMGSVRSVEQFHKYLPMFLNDRP 1020
DB 961 LPPSSCCRLYISGPNKDKECPSTVNSLNCJKNCSITMGSVRSVEQFHKYLPMFLNDRP 1020
QY 1021 NIKCPKGLAAVSTSVNLTSQGVL-----ASRFMAVH 1053
DB 1021 NIKCPKGLAAVSTSVNLTSQGVL-----ASRFMAVH 1053
QY 1054 KPLKNSQDYTEALRAARELANITADLRKVPCTDPAPFEPVPTITVNFYEQYITLPEGL 1113
DB 1054 KPLKNSQDYTEALRAARELANITADLRKVPCTDPAPFEPVPTITVNFYEQYITLPEGL 1113
QY 1081 KPLKNSQDYTEALRAARELANITADLRKVPCTDPAPFEPVPTITVNFYEQYITLPEGL 1140
DB 1081 KPLKNSQDYTEALRAARELANITADLRKVPCTDPAPFEPVPTITVNFYEQYITLPEGL 1140
QY 1114 FMLSCLVPTFAVSCLLIGLDSGLNLNLSIYMIIVDTGFPALMDISYNAVSLNLS 1173
DB 1114 FMLSCLVPTFAVSCLLIGLDSGLNLNLSIYMIIVDTGFPALMDISYNAVSLNLS 1173
QY 1141 FMLSCLVPTFAVSCLLIGLDSGLNLNLSIYMIIVDTGFPALMDISYNAVSLNLS 1200
DB 1141 FMLSCLVPTFAVSCLLIGLDSGLNLNLSIYMIIVDTGFPALMDISYNAVSLNLS 1200
QY 1174 AVGMSVEPVSHITRSFAISTKPTWLBRAKEATISMSAVFAGVAMTNLPGLIYLGLAKQ 1233
DB 1174 AVGMSVEPVSHITRSFAISTKPTWLBRAKEATISMSAVFAGVAMTNLPGLIYLGLAKQ 1233
QY 1201 AVGMSVEPVSHITRSFAISTKPTWLBRAKEATISMSAVFAGVAMTNLPGLIYLGLAKQ 1260
DB 1201 AVGMSVEPVSHITRSFAISTKPTWLBRAKEATISMSAVFAGVAMTNLPGLIYLGLAKQ 1260
QY 1234 LIDIFFFRNLMLTLTGLHGLVFLPYIILSYVGPDPVPALALQKAAEEVAAMVMAVASC 1293
DB 1234 LIDIFFFRNLMLTLTGLHGLVFLPYIILSYVGPDPVPALALQKAAEEVAAMVMAVASC 1293
QY 1261 LIDIFFFRNLMLTLTGLHGLVFLPYIILSYVGPDPVPALALQKAAEEVAAMVMAVASC 1320
DB 1261 LIDIFFFRNLMLTLTGLHGLVFLPYIILSYVGPDPVPALALQKAAEEVAAMVMAVASC 1320
QY 1294 NHPRSVSTADNIYVNSFEGSIRKAGALISNPLFNNGRQF 1332
DB 1294 NHPRSVSTADNIYVNSFEGSIRKAGALISNPLFNNGRQF 1332
QY 1321 NHPRSVSTADNIYVNSFEGSIRKAGALISNPLFNNGRQF 1359
DB 1321 NHPRSVSTADNIYVNSFEGSIRKAGALISNPLFNNGRQF 1359

```

```

RESULT 10
ABG22693
ID ABG22693 standard; protein; 1344 AA.
AC ABG22693;
XX
DT 18-FEB-2002 (first entry)

```

```

XX DE Novel human diagnostic protein #22684.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN MO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US008631.
XX PR 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Dmanac RT, Liu C, Tang YT;
XX DR N-PSDB; AAS86880.
XX DR WPI: 2001-639362/73.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.
XX PS Claim 20; SEQ ID NO 53052; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX CC sequences. (I) is useful as hybridisation probe, polymerase chain
XX CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX CC and in recombinant production of (II). The polynucleotides are also used
XX CC in diagnostics as expressed sequence tags for identifying expressed
XX CC genes. (I) is useful in gene therapy techniques to restore normal
XX CC activity of (II) or to treat disease states involving (II). (II) is
XX CC useful for generating antibodies against it, detecting or quantitating a
XX CC polypeptide in tissue, as molecular weight markers and as a food
XX CC supplement. (II) and its binding partners are useful in medical imaging
XX CC of sites expressing (II). (I) and (II) are useful for treating disorders
XX CC involving aberrant protein expression or biological activity. The
XX CC polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
XX CC amino acid sequences of the invention. Note: The sequence data for this
XX CC patent did not appear in the printed specification, but was obtained in
XX CC electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 1344 AA;
QY Query Match 94.6%; Score 6536; DB 4; Length 1344;
QY Best local similarity 96.7%; Pred. No. 0;
QY Matches 1271; Conservative 7; Mismatches 21; Indels 16; Gaps 4;
DB 1 MAEAGLKGWLLMALLRLAQSSEPYTTIHQPGYCAFYDECGKNPELSGSLMTLSNVSCLSN 60
DB 1 MAEAGLKGWLLMALLRLAQSSEPYTTIHQPGYCAFYDECGKNPELSGSLMTLSNVSCLSN 60
QY 61 TPARKITGDHLILLOKICPRLYTGPTQACCSAKQVLSLEASTLTKALLTRCPACSDNF 120
DB 61 TPARKITGDHLILLOKICPRLYTGPTQACCSAKQVLSLEASTLTKALLTRCPACSDNF 120
QY 121 VNLHCNNTCSPNQSLFINVTRVAQLGAGQLPAVVAVEAFYQHSFABQSYDSCSRVAVPAA 180
DB 121 VNLHCNNTCSPNQSLFINVTRVAQLGAGQLPAVVAVEAFYQHSFABQSYDSCSRVAVPAA 180
QY 181 ATLAAGTMCVYGSAICNARWLNFQGTGNGLAPLDITFHLLPEQAVGSGIOPLANEV 240
DB 181 ATLAAGTMCVYGSAICNARWLNFQGTGNGLAPLDITFHLLPEQAVGSGIOPLANEV 240

```

Db 181 ATLAVGTCGYGSLCNAQRWLNQGTGNGLAPLDITFHLBEGQAVSGSIQPLNBSV 240
Qy 241 ARCNESQGDVATCSCODCAASC- -A1ARPOLDSTFYAGMPGSLVL11ILGSAFAY 298
Db 241 ARCNESQGDVATCSCODCAASCPRAPRSTPSTW--RQMPGSLVL11ILGSAFAY 298
Qy 299 TLLVGFVAPARADSKMVDPKKGTSLSDKLSFSTHTLLGQFQCGKGTWASMPRLTILVL 358
Db 299 TLLVGFVAPARADSKMVDPKKGTSLSDKLSFSTHTLLGQFQCGKGTWASMPRLTILVL 358
Qy 359 SVTPVVALAAGLVFELTTPDPELMSAPNSQARSEKAFHDQGFPEFRINOYLITAPRNS 418
Db 359 SVTPVVALAAGLVFELTTPDPELMSAPNSQARSEKAFHDQGFPEFRINOYLITAPRNS 418
Qy 419 SYRYDSSLGPPNFGSILDLDBLLELLEQLERLRLQVMSPEAQRNLSQDICYAPLND 478
Db 419 SYRYDSSLGPPNFGSILDLDBLLELLEQLERLRLQVMSPEAQRNLSQDICYAPLND 478
Qy 479 NTSLYDCCINSLLQYFQNNRTLLLTANQTLMGQTSQVDWKDHFLYCANAPLTFKQDTNL 538
Db 479 NTSLYDCCINSLLQYFQNNRTLLLTANQTLMGQTSQVDWKDHFLYCANAPLTFKQDTNL 538
Qy 539 ALSGADYGAPEPFLAIGYKGYKDYSEAEALIMFESLNNYPADGPRLAQAALMEBAFLE 598
Db 539 ALSGADYGAPEPFLAIGYKGYKDYSEAEALIMFESLNNYPADGPRLAQAALMEBAFLE 598
Qy 599 EMRAFORRMAGMFOVYFTAERSLDEINRTAEDLP1FATSYIVIFLYISLALGSYSWS 658
Db 599 EMRAFORRMAGMFOVYFTAERSLDEINRTAEDLP1FATSYIVIFLYISLALGSYSWS 658
Qy 659 RVMVDSKATLGLGVAIVIGAVMAAMGFFSYGIRSSLVITLOVPEPLVSVGADNIFTV 718
Db 659 RVMVDSKATLGLGVAIVIGAVMAAMGFFSYGIRSSLVITLOVPEPLVSVGADNIFTV 718
Qy 719 LEYORLPRRPGPREVHIGRALGRVAPSWMLCSLSRATCEFLGALTPMAVATPALTSGL 778
Db 719 LEYORLPRRPGPREVHIGRALGRVAPSWMLCSLSRATCEFLGALTPMAVATPALTSGL 778
Qy 779 AVILDFLLQMSAFVALLSLDSKROEASRLDVCVCVPOELPPPGQEGILLGFPKAVAP 838
Db 779 AVILDFLLQMSAFVALLSLDSKROEASRLDVCVCVPOELPPPGQEGILLGFPKAVAP 838
Qy 839 FLIHMITRGV---LLFLFALPGVSLVSMCHI SVGLDOELALPKDSYLLDYFLPLNRYE 894
Db 839 FLIHMITRGV---LLFLFALPGVSLVSMCHI SVGLDOELALPKDSYLLDYFLPLNRYE 894
Qy 895 EVGAIVYFTTTLGVNPSSEAGMNAICSSAGCNPFSTOKIOVATEPEBOSYLAIPASSAV 954
Db 895 EVGAIVYFTTTLGVNPSSEAGMNAICSSAGCNPFSTOKIOVATEPEBOSYLAIPASSAV 954
Qy 955 DDFIDMLTPSSCCRLYISGPNKDKCPSTVNSLNCILKNCMSITMGSVRSVEQFHKYLPW 1014
Db 955 DDFIDMLTPSSCCRLYISGPNKDKCPSTVNSLNCILKNCMSITMGSVRSVEQFHKYLPW 1014
Qy 1015 FLNDRPNIKCPKGLAAVSTVNLTSQGVTLASRFMAVHKPKNSODYTEALRAABELAA 1074
Db 1015 FLNDRPNIKCPKGLAAVSTVNLTSQGVTLASRFMAVHKPKNSODYTEALRAABELAA 1074
Qy 1075 NITLALRKVPGDPAPEVEPPYITITVYFROYLTLLPEGLFMLSLCIVPFAVASCILLGD 1134
Db 1075 NITLALRKVPGDPAPEVEPPYITITVYFROYLTLLPEGLFMLSLCIVPFAVASCILLGD 1134
Qy 1135 LRSGLNLNLSTVMIIVDTVGFNALMDISYNAVSLINLVASVMSVEFVSHITRSFAISRK 1194
Db 1135 LRSGLNLNLSTVMIIVDTVGFNALMDISYNAVSLINLVASVMSVEFVSHITRSFAISRK 1194
Qy 1195 PTWLERAKAATISMSGSAVAGVAMTNLPGLIVLGLAKAQLIOI FFERLNLITLLGLLHG 1254
Db 1195 PTWLERAKAATISMSGSAVAGVAMTNLPGLIVLGLAKAQLIOI FFERLNLITLLGLLHG 1254
Qy 1255 LVFLVLIISYVGPDPNPALALEOKABEAA-----VAANVVASCPNHPSPVST 1301
Db 1255 LVFLVLIISYVGPDPNPALALEOKABEAA-----VAANVVASCPNHPSPVST 1301
Qy 1301 LVFLVLIISYVGPDPNPALALEOKABEAA-----VAANVVASCPNHPSPVST 1301
Db 1301 LVFLVLIISYVGPDPNPALALEOKABEAA-----VAANVVASCPNHPSPVST 1301

RESULT 11
ADJ27295
ID ADJ27295 standard; protein; 1331 AA.
XX
AC ADJ27295;
XX
DT 20-MAY-2004 (first entry)
XX
DE Rat NPC1L1.
XX
KW Niemann-Pick disease; type C1; gene-like 1; NPC1L1; trans-golgi network;
KW plasma membrane; transport signal; promoter;
KW sterol regulated element binding protein 1; SREBP1;
KW binding consensus sequence; transmembrane domain; sterol-sensing domain;
KW SSD; cholesterol; NPC1; receptor; Niemann-Pick C1 disease; intestinal;
KW cholesterol absorption; serum cholesterol; hyperlipidaemia;
KW atherosclerosis; coronary heart disease; stroke; arteriosclerosis.
XX
OS Rattus sp.
XX
PN MO2004009772-A2.
XX
PD 29-JAN-2004.
XX
PF 17-JUL-2003; 2003WO-US022467.
XX
PR 19-JUL-2002; 2002US-0397442P.
XX
PA (SCHE) SCHERING CORP.
XX
PI Altmann SW, Murgolo NJ, Wang LQ, Graziano MP;
XX
DR MPI: 2004-132945/13.
XX
DR N-PSDB: ADJ27294.
XX
PT New Niemann-Pick disease, type C1, gene-like 1 polypeptide, for detecting
PT or identifying antagonists of NPC1L1 for inhibiting intestinal
PT cholesterol absorption in a subject, or for treating elevated serum
PT cholesterol or stroke.
XX
PS Claim 1; SEQ ID NO 2; 125bp; English.
XX
CC This sequence represents a Niemann-Pick disease, type C1, gene-like 1
CC (NPC1L1) polypeptide. NPC1L1 is an N-glycosylated protein which contains
CC a motif ADJ27331 which acts as a trans-golgi network to plasma membrane
CC transport signal, and which exhibits limited tissue distribution and
CC gastrointestinal abundance. The human NPC1L1 promoter sequence contains a
CC sterol regulated element binding protein 1 (SREBP1) binding consensus
CC sequence. NPC1L1 has 13 transmembrane spanning segments and a sterol-
CC sensing domain (SSD) which is involved in sensing cholesterol levels.
CC possibly by a mechanism which involves direct cholesterol binding. NPC1L1
CC has 428 amino acid sequence homology to human NPC1, a receptor
CC responsible for Niemann-Pick C1 disease. The NPC1L1 polypeptides or
CC polynucleotides are useful for detecting or identifying antagonists of
CC NPC1L1, which can be used for inhibiting intestinal cholesterol
CC absorption in a subject, or for treating medical conditions including
CC elevated serum cholesterol, hyperlipidaemia, arteriosclerosis, coronary
CC heart disease, stroke or arteriosclerosis.
XX
SQ Sequence 1331 AA;
XX
Query Match 78.5%; Score 5421.5; DB 8; Length 1331;
Best Local Similarity 77.8%; Pred. No. 0;
Matches 1039; Conservative 120; Mismatches 165; Indels 7; Gaps 6;
Qy 1 MAEAGLRGMLMALRLAAGEPYTHQPGYCAFYDECGNPKELSGMLTSLNVSCLSN 60
Db 1 MAEAGLRGMLMALRLAAGEPYTHQPGYCAFYDECGNPKELSGMLTSLNVSCLSN 59
Qy 61 TPARKITGDHLLLOKITEPRITGPNTQ-ACCSAKQVLSLASLSTTALLTRCAGSDN 119
Db 61 TPARKITGDHLLLOKITEPRITGPNTQ-ACCSAKQVLSLASLSTTALLTRCAGSDN 119

Db 60 TPAAHVTGHEHLALLOBIICBPLVNGPNTTACSTKQLLSLESSMSITTKALLTRCPACSDN 119
 QY 120 FVNLHCNTCSPNQSLFINVTRVAOLGAGQLPAVVAVEAFYOHSPAEOSYDSCSRVRPA 179
 Db 120 FVSLHCNTCSPNQSLFINVTRVERKAGBPAAVVAVEAFYOHSPAEKABESSQVRIPA 179
 QY 180 AATLAVGTMGVYGSALCNANQWLNFGDGTNGIAPLDITPHLLEBQAVGSGIQPLNBS 239
 Db 180 AASLAVGSMGVYGSALCNANQWLNFGDGTNGIAPLDITPHLLEBQALPDGIQPLNGK 239
 QY 240 VARCNESQGDVATCGCCODCASCPAIARPQALDSTFYIGOMGSLVLIILCSVAVAT 299
 Db 240 IAPCNESQGDVATCGCCODCASCPVIPPPEALRPFGYGRMPGWLALIIIFPAVFLVS 299
 QY 300 ILLVGRVAPARSKSMVDPKKGTSLSDKLSFSTHTLLGQFQGMGTWVASWPLTIVLS 359
 Db 300 AVLVRLRVVSNRKNKAEBQEPKLPKHKLSPHIILGFFPNMGTRVASMPLIVLALS 359
 QY 360 VTPVVALAAGLVETELTDPVELMSAPNSQARSEKAFHDOHFGPFPRITNOVILTAENRS 419
 Db 360 FIVVIALAAGLVETELTDPVELMSAPKSGARKEKSFHDEHFGPFPRITNOIPTVAENRS 419
 QY 420 YKXDSLILPKNFSGLIDDLLELLELQERLHLQVSPBQARISLQDICTAPLNPDN 479
 Db 420 YKXDSLILKSNFSGLISLDFLELLELQERLHLQVSPBQARISLQDICTAPLNPVN 479
 QY 480 TSLYDCINSLQYFQNNRTLLLTANQTLMGQTSQVMDHLYCANAPLTFKDGTA 539
 Db 480 TSLSDCCVNSLQYFQNNRTLLMLTANQTLMGQTSQVMDHLYCANAPLTFKDGISLA 539
 QY 540 LSCMDYGAPEVPEFLAIGYKGDYSEAEALIMTFSLNNYPADPRLAQAQKWEAEFLER 599
 Db 540 LSCMDYGAPEVPEFLAVGQYQGDYSEAEALITTFSLNNYPADPRLAQAQKWEAEFLER 599
 QY 600 MEAFQRMAGMFOVTFARSLDEINRTTADLPPIATSYIVIFYIYISIALGYSYSWSR 659
 Db 600 MESFOGNTSDKQVAFSAERSLEDEINRTTIQDLPVFAVSIIIVFLYISIALGYSYRCSR 659
 QY 660 VAVDSATTLGLGVANVLGAVMAAMGFYSYIGRSSVLVLOVVPFLVLSGAMNITFVL 719
 Db 660 VAVDSATTLGLGVAVVLGAVLAAVMAAMGFYSYIGRSSVLVLOVVPFLVLAAGANITFVL 719
 QY 720 EYQRLPRRGEPEPEVAHIGRALGRVAPSMILCSLSEAIICEFLGALTMPAVRTALTSGLA 779
 Db 720 EYQRLPRMGEQGEAHIGRTLSVAPSMILCSLSEAIICEFLGALTMPAVRTALTSGLA 779
 QY 780 VILDFLLQNSAFVALLSLDSKQGEASRLDYCCCVKQDELPPQGGEGLLGFFQKAYAP 839
 Db 780 IILDFLLQNTAFVALLSLDSKQGEASRPDVLCCFSTRKLPPEPKKEGALLRFRKIYAP 839
 QY 840 LHMVTRGVVLLFLALFGVSLYSMGHISVGLDQELALPRDSYLLDYFLFANFFVGVAP 899
 Db 840 LHMFRPVPVMLLFLFLGANLYLMCNVINGLDQELALPRDSYLLDYFLFANFFVGVAP 899
 QY 900 VYFVTLTYGVNFSSEAGMNAICSSAGCNFSFTOKIOVATPEFPOSYIATIPASSWVDIF 959
 Db 900 VYFVTLTYGVNFSSEAGMNAICSSAGCKSBSLTOKIOVASFPQSYIATIPASSWVDIF 959
 QY 960 WLP-SSCRLYISGNKDKFCPSTYNSLNCNKSITMGVSRVSVGEFHXYLPWFLND 1018
 Db 960 WLPSSSCRLYIRGPHKDEFCSTDTSEFCLNCKMNRITGPVRPAEQGHXYLPWFLND 1019
 QY 1019 RNNIKPKGGLAAYSTVNLTSPOVLASRFMAVHKPLKNSQOYTERALRAELANITR 1078
 Db 1020 PPPIRCPKGLAAYRSTVNLSSDGOVIASQFMAVHPRLRNSODFTELRARSRLANITR 1079
 QY 1079 DLKRVGTDPAPFVPPYTTINVFYBOYLTLPREGLEFMLSCLVPTFVASCILGLDLRS 1138
 Db 1080 DLKRVGTDPBNFVPPYTTISNVFYQYLYLPBGIFITLALCFPTFVVCILLGLDMCSG 1139
 QY 1139 LNLLSIVMLVDTVGFMAIMDISYNAVSLINLVSAGVSVFVSHITRSFAISTKPTWL 1198
 Db 1140 INLLSIIMILVDITIGLMAVWGISYNAVSLINLVTAVGMSVERVSHITRSFAVSTKPTRL 1199

QY 1199 ERAKEATISMSAVFAGVAMTNLPGILVGLAKAQLIQIFPRLNLLITLLGLHGLVPL 1258
 Db 1200 ERAKQATVMSGAVSFAGVAMTNFPGILILGFAQAQLIQIFPRLNLLITLLGLHGLVPL 1259
 QY 1259 PVLSTYVGPVDVQALBQGRABEAAVAAVMVASCPNHPBSRVSTADNIIYNHSPBSGI-KG 1317
 Db 1260 PVLSTYLGPDVQALVQEEKLASEA-AVAPBESCPQYPSPADADAN--VNYGPAPELAWG 1316
 QY 1318 AGAISNFPENNRQF 1332
 Db 1317 ANAARSLPKSDQK 1331
 RESULT 12
 ADY60738
 ID ADY60738 standard; protein; 1331 AA.
 XX
 AC ADY60738;
 XX
 DT 19-MAY-2005 (first entry)
 XX
 DE Rat Niemann-Pick C1-like protein 1 (NPC1L1) protein.
 XX
 KW Niemann-Pick C1-like protein 1; Antilipemic; Antiarteriosclerotic;
 KW Cardiant; Vasotropic; Cerebroprotective; cholesterol; hyperlipidemia;
 KW metabolic disorder; atherosclerosis; cardiovascular disease;
 KW coronary artery disease; cerebrovascular ischemia; neurological disease;
 KW arteriosclerosis; transgenic animal.
 XX
 OS Rattus sp.
 XX
 PN W02005015988-A1.
 XX
 PD 24-FEB-2005.
 XX
 PF 16-DEC-2003; 2003WC-US040113.
 XX
 PR 17-JUL-2003; 2003US-00621758.
 PR 22-AUG-2003; 2003US-00646301.
 PR 16-SEP-2003; 2003US-00663208.
 XX
 PA (SCHE) SCHERING CORP.
 XX
 PI Altmann SM, Murgolo NJ, Wang LQ, Graziano MP;
 XX
 DR MPI; 2005-284403/29.
 XX
 DR N-PSDB; ADY60737, ADY60746.
 XX
 PT Novel isolated NPC1L1 polypeptide, useful for identifying antagonist of
 PT NPC1L1 that inhibits NPC1L1 mediated sterol or 5 alpha-sterol uptake, for
 PT treating hyperlipidemia or atherosclerosis in mouse, rat or human.
 XX
 PS Claim 30; SEQ ID NO 2; 146p; English.
 XX
 CC The invention relates to an isolated Niemann-Pick C1-like protein 1
 CC (NPC1L1) polypeptide comprising 42 or more contiguous amino acids from an
 CC amino acid sequence appearing as ADY60738,ADY60740 or ADY60748 (rat,
 CC human and mouse NPC1L1). Also included are an isolated polynucleotide
 CC encoding NPC1L1 (appearing as ADY60737,ADY60739 or ADY60747), a
 CC recombinant vector comprising the polynucleotide, a host cell comprising
 CC the vector, an isolated antibody which specifically binds to NPC1L1 (or
 CC to a rat NPC1L1 peptide sequence appearing as ADY60775-ADY60778),
 CC preparation of NPC1L1, a mutant transgenic mouse comprising a homozygous
 CC mutation of endogenous chromosomal NPC1L1 (where the mouse does not
 CC produce any functional NPC1L1 protein), an offspring or progeny of the
 CC mouse having inherited a mutated NPC1L1 allele of the mouse, screening a
 CC sample for an intestinal sterol or Salpha-sterol absorption antagonist,
 CC inhibiting NPC1L1 mediated sterol or Salpha-sterol uptake in a subject, a
 CC kit (comprising ezetimibe or substituted azetidione in a pharmaceutical
 CC dosage form, and information indicating that NPC1L1 is a target of
 CC ezetimibe or substituted azetidione), decreasing the level of intestinal
 CC sterol or Salpha-sterol absorption in a subject (involving reducing the

PT glucuronide and a candidate compound and determining if compound binds to human NPC1L1.

XX Example 2; SEQ ID NO 2; 215pp; English.

XX The invention relates to identifying a ligand of NPC1L1. The method involves contacting human NPC1L1 with a detectably labeled substituted 2-azetidineone glucuronide and a candidate compound and determining whether the candidate compound binds to human NPC1L1. In identifying a ligand of NPC1L1, the KD value is 100 nM or lower, preferably 50 nM, 20 nM, or 10 nM or lower. The detectably labeled substituted 2-azetidineone glucuronide is 3-5 labeled compound 2, given in the specification. NPC1L1 ligands are useful for stimulating or blocking the activity of NPC1L1, and for creating conditions caused or mediated by NPC1L1. It is useful for reducing the incidence of hyperlipidemia, atherosclerosis, coronary heart disease, stroke, or arteriosclerosis. The present sequence represents a rat NPC1L1, a N-glycosylated protein.

XX Sequence 1331 AA;

Query Match 78.5%; Score 5421.5; DB 9; Length 1331;

Best Local Similarity 77.8%; Pred. No. 0;

Matches 1039; Conservative 120; Mismatches 169; Indels 7; Gaps 6;

QY 1 MBEAGLGMWLLWALLRLAQSEPTTTHQPGYCAFYDECGKNBELSGMTLSNVSCLSN 60
DB 1 MAAAWL-GWLLWALLLSAAQGLYTPKHEAGVCTFEECCGNBELSGGLTSLNVSCLSN 59
QY 61 TPARKTGHLLLOKICRRLTYGENTQ-ACCSAKOLVLSBASLSTTKALTRCPACSDN 119
DB 60 TPARKTGHLLLOKICRRLTYGENTQ-ACCSAKOLVLSBASLSTTKALTRCPACSDN 119
QY 120 PNLHCHNTCSPNQSFINTVRAOAGOLPAVNVYEAFOYHSFAEOSYDCSRARVA 179
DB 120 PNLHCHNTCSPNQSFINTVRAOAGOLPAVNVYEAFOYHSFAEOSYDCSRARVA 179
QY 120 PNLHCHNTCSPNQSFINTVRAOAGOLPAVNVYEAFOYHSFAEOSYDCSRARVA 179
DB 120 PNLHCHNTCSPNQSFINTVRAOAGOLPAVNVYEAFOYHSFAEOSYDCSRARVA 179
QY 180 AATLAVGTCGVYGSALCNARWLNFGDGTNGLAFLDITFHLLBEGAVSGGIQPLNBG 239
DB 180 AATLAVGTCGVYGSALCNARWLNFGDGTNGLAFLDITFHLLBEGAVSGGIQPLNBG 239
QY 180 AATLAVGTCGVYGSALCNARWLNFGDGTNGLAFLDITFHLLBEGAVSGGIQPLNBG 239
DB 180 AATLAVGTCGVYGSALCNARWLNFGDGTNGLAFLDITFHLLBEGAVSGGIQPLNBG 239
QY 240 VARCNESQDDVATCGCCDCAASCRAIARPOLDSTFYIGQMGSVLIIILCSVAVVT 299
DB 240 VARCNESQDDVATCGCCDCAASCRAIARPOLDSTFYIGQMGSVLIIILCSVAVVT 299
QY 240 VARCNESQDDVATCGCCDCAASCRAIARPOLDSTFYIGQMGSVLIIILCSVAVVT 299
DB 240 VARCNESQDDVATCGCCDCAASCRAIARPOLDSTFYIGQMGSVLIIILCSVAVVT 299
QY 300 ILLVGRVAPARSKMVPKGTSLSDKLSFSTHTLLGQFQCGWGTWVSWPLTIVLS 359
DB 300 ILLVGRVAPARSKMVPKGTSLSDKLSFSTHTLLGQFQCGWGTWVSWPLTIVLS 359
QY 360 VIIPVVALAGLVTELTTPDVELMSAPNSQARSEKAFHQHFGPFRITNOVILTAENRS 419
DB 360 VIIPVVALAGLVTELTTPDVELMSAPNSQARSEKAFHQHFGPFRITNOVILTAENRS 419
QY 420 YRYSILLGPKNFGSGLIDDLLELLELORLHLQVWSPBEAQRNLSLODICTAIPNPN 479
DB 420 YRYSILLGPKNFGSGLIDDLLELLELORLHLQVWSPBEAQRNLSLODICTAIPNPN 479
QY 480 TSLYDCINSLQYQPNNTFLILLTNQTLMGOTQVDMQDHLXANAPLTFKDDTALA 539
DB 480 TSLYDCINSLQYQPNNTFLILLTNQTLMGOTQVDMQDHLXANAPLTFKDDTALA 539
QY 480 TSLYDCINSLQYQPNNTFLILLTNQTLMGOTQVDMQDHLXANAPLTFKDDTALA 539
DB 480 TSLYDCINSLQYQPNNTFLILLTNQTLMGOTQVDMQDHLXANAPLTFKDDTALA 539
QY 540 LSCMADYGAAPVFPFLAIGYKGYSEAEALINTFSLNNYPADQPRLAQKMEAEFLBE 599
DB 540 LSCMADYGAAPVFPFLAIGYKGYSEAEALINTFSLNNYPADQPRLAQKMEAEFLBE 599
QY 540 LSCMADYGAAPVFPFLAIGYKGYSEAEALINTFSLNNYPADQPRLAQKMEAEFLBE 599
DB 540 LSCMADYGAAPVFPFLAIGYKGYSEAEALINTFSLNNYPADQPRLAQKMEAEFLBE 599
QY 600 MRAFORMAGMFOVTTAERSLDEINRTAEDLPFANSYIYIPLYISIALGYSYSWSR 659
DB 600 MRAFORMAGMFOVTTAERSLDEINRTAEDLPFANSYIYIPLYISIALGYSYSWSR 659
QY 600 MRAFORMAGMFOVTTAERSLDEINRTAEDLPFANSYIYIPLYISIALGYSYSWSR 659
DB 600 MRAFORMAGMFOVTTAERSLDEINRTAEDLPFANSYIYIPLYISIALGYSYSWSR 659
QY 660 VMDSDKATLGLGAVAVILGAVVAMAGFESYLGRSSIVILQVAPLVLISGANITFLV 719
DB 660 VMDSDKATLGLGAVAVILGAVVAMAGFESYLGRSSIVILQVAPLVLISGANITFLV 719
QY 660 VMDSDKATLGLGAVAVILGAVVAMAGFESYLGRSSIVILQVAPLVLISGANITFLV 719
DB 660 VMDSDKATLGLGAVAVILGAVVAMAGFESYLGRSSIVILQVAPLVLISGANITFLV 719
QY 720 EYORLPPEGEQREAHIGRTIGSVAPSMILCSLSSEALICFFDALTPMPAVRTFALTSGLA 779
DB 720 EYORLPPEGEQREAHIGRTIGSVAPSMILCSLSSEALICFFDALTPMPAVRTFALTSGLA 779

DB 720 EYORLPPEGEQREAHIGRTIGSVAPSMILCSLSSEALICFFDALTPMPAVRTFALTSGLA 779
QY 780 VILDFLQMSAFVALLSDSKROEASRLDVCCCVARPEPPQEGSLAGFORAVAP 839
DB 780 VILDFLQMSAFVALLSDSKROEASRLDVCCCVARPEPPQEGSLAGFORAVAP 839
QY 840 LHMTRGVALLFLALPGVSLYSMCHISVGLDQELAPKDSYLLDYFLNRYEVCAP 899
DB 840 LHMTRGVALLFLALPGVSLYSMCHISVGLDQELAPKDSYLLDYFLNRYEVCAP 899
QY 900 VYFVTTLGNSESEAGMALICSSACNNPSFQXQYOYTEPEQSYLAIIPASSWDDFD 959
DB 900 VYFVTTLGNSESEAGMALICSSACNNPSFQXQYOYTEPEQSYLAIIPASSWDDFD 959
QY 960 WLTP-SSCCRLYISGPNKDKFCPTVNSLNCCKNGSITMSGVSREVEQFHLYLPFLND 1018
DB 960 WLTP-SSCCRLYISGPNKDKFCPTVNSLNCCKNGSITMSGVSREVEQFHLYLPFLND 1018
QY 960 WLTP-SSCCRLYISGPNKDKFCPTVNSLNCCKNGSITMSGVSREVEQFHLYLPFLND 1018
DB 960 WLTP-SSCCRLYISGPNKDKFCPTVNSLNCCKNGSITMSGVSREVEQFHLYLPFLND 1018
QY 1019 RPNKCPKGLAAVSTVNLTSQGVLASRFMAVHKPLKNSQDYTEALRAABELANITA 1078
DB 1019 RPNKCPKGLAAVSTVNLTSQGVLASRFMAVHKPLKNSQDYTEALRAABELANITA 1078
QY 1079 DLRKTPGTDPAEVPYTTITNVFYEOYLTLPEGLPMLSLCVPTFPANSCLLGLDRSG 1138
DB 1079 DLRKTPGTDPAEVPYTTITNVFYEOYLTLPEGLPMLSLCVPTFPANSCLLGLDRSG 1138
QY 1139 LNTLSIWMILVDYTGFMALMDISYNAVSLINLVASGVSVFVSHITRSPASTKPTWL 1198
DB 1139 LNTLSIWMILVDYTGFMALMDISYNAVSLINLVASGVSVFVSHITRSPASTKPTWL 1198
QY 1199 ERAKEATISMGSAVPAVAMTNLPGILVLGLAKAOLIOIFERRMLITLLGLHGLVFL 1258
DB 1199 ERAKEATISMGSAVPAVAMTNLPGILVLGLAKAOLIOIFERRMLITLLGLHGLVFL 1258
QY 1259 PYLTSYGPDPVAPALAEQKRAEVAAMVASCNHRSTANDIYNHSEFEST-KG 1317
DB 1259 PYLTSYGPDPVAPALAEQKRAEVAAMVASCNHRSTANDIYNHSEFEST-KG 1317
QY 1318 AGAISNPLPNNGROF 1332
DB 1318 AGAISNPLPNNGROF 1332
QY 1317 ANAARSLPKSDQKF 1331
DB 1317 ANAARSLPKSDQKF 1331
RESULT 14
ADJ27305
ID ADJ27305 standard; protein; 1333 AA.
XX AC ADJ27305;
XX DT 20-MAY-2004 (first entry)
XX DE Mouse NPC1L1.
XX NI Niemann-Pick disease; type C1; gene-like 1; NPC1L1; trans-golgi network;
XX PL plasma membrane; transport signal; promoter;
XX ST sterol regulated element binding protein 1; SREBP1;
XX BI binding consensus sequence; transmembrane domain; sterol-sensing domain;
XX SS SCD; cholesterol; NPC1; receptor; Niemann-Pick C1 disease; intestinal;
XX KW cholesterol absorption; serum cholesterol; hyperlipidemia;
XX KW atherosclerosis; coronary heart disease; stroke; arteriosclerosis.
OS Mus sp.
XX WO2004009772-A2.
XX PD 29-JAN-2004.
XX PF 17-JUL-2003; 2003WO-US022467.
XX PR 19-JUL-2002; 2002US-0397442P.
XX PA (SCHER) SCHERING CORP.

XX Altman SW, Murgolo NJ, Wang LQ, Graziano MP;
 XX MPI; 2004-132945/13.
 DR N-PSDB; ADJ27304.
 XX
 PT New Niemann-Pick disease, type C1, gene-like 1 polypeptide, for detecting
 PT or identifying antagonists of NPC1L1 for inhibiting intestinal
 PT cholesterol absorption in a subject, or for treating elevated serum
 PT cholesterol or stroke.
 XX
 PS Claim 1; SEQ ID NO 12; 125pp; English.
 XX
 CC This sequence represents a Niemann-Pick disease, type C1, gene-like 1
 CC (NPC1L1) polypeptide. NPC1L1 is an N-glycosylated protein which contains
 CC a motif ADJ27331, which acts as a trans-golgi network to plasma membrane
 CC transport signal, and which exhibits limited tissue distribution and
 CC gastrointestinal abundance. The human NPC1L1 promoter sequence contains a
 CC sterol regulated element binding protein 1 (SREBP1) binding consensus
 CC sequence. NPC1L1 has 13 transmembrane spanning segments and a sterol-
 CC sensing domain (SSD) which is involved in sensing cholesterol levels,
 CC possibly by a mechanism which involves direct cholesterol bonding. NPC1L1
 CC has 428 amino acid sequence homology to human NPC1, a receptor
 CC responsible for Niemann-Pick C1 disease. The NPC1L1 polypeptides or
 CC polynucleotides are useful for detecting or identifying antagonists of
 CC NPC1L1, which can be used for inhibiting intestinal cholesterol
 CC absorption in a subject, or for treating medical conditions including
 CC elevated serum cholesterol, hyperlipidaemia, atherosclerosis, coronary
 CC heart disease, stroke or arteriosclerosis.
 CC
 XX Sequence 1333 AA;
 SQ

Query Match 78.3%; Score 5407; DB 8; Length 1333;
 Best Local Similarity 77.2%; Pred. No. 0;
 Matches 1028; Conservative 126; Mismatches 174; Indels 4; Gaps 4;

QY 4 AGRGRLMALILRLAOSPPYTTIHOPGYCAFDCEGKPELSGSLMTSNISCLSNTPA 63
 DB 3 AAMQGLMALILNLSOGELVTPTHKAGCTPYEEGKNPELSGGTISLNSISCLSNTPA 62
 QY 64 RITGDHLLLOKICRKYLTGPN-TQACSAKQVLSEASLSTTKALLRCPACSNPNFN 122
 DB 63 RHVTGDHLLLOKICRKYLTGPN-TQACSAKQVLSEASLSTTKALLRCPACSNPNFN 122
 QY 121 IHCNHTCSPNOSLFINVTVAQLGAGQLPAVVAEYAFYQHSFAEOSYDSCSRVPAAPAT 182
 DB 123 IHCNHTCSPDQSLFIVTRVVGQDPQLPAVVAEYAFYQHSFAEKAYECSVRIPAAAS 182
 QY 183 LAVGTCGYVGSALCNAQWMLNFOGDTGNGLAFLDITPHLLBFGQAVSGGIQPLNEGVAR 242
 DB 183 LAVGTCGYVGSALCNAQWMLNFOGDTGNGLAFLDITPHLLBFGQAVSGGIQPLNEGVAR 242
 QY 243 CNEGSGDVATCSGCCDCAASCPAIAIRPOLDSFTVYLGOMGSLVLIILICSPAVAVTIL 302
 DB 243 CNEGSGDSDAACSQCDCASCPVIPPALRPSFTYGRMPGMALIIITTAFAVLISVLV 302
 QY 303 VGFRTVAARDKSKMVDPKKGTSLSDXLSEFTHTLLGQFQGWGTWVASWPLTILVLSVIP 362
 DB 303 VLRVAASNKKTKTSGSGAPNLPRKRRTSPHTVLGRFPESNGTRVASFVTLVLSFTV 362
 QY 363 VVLAAGLVFTLTDVPELMSAPNSQARSEKAFHDQHPFFRINQVILTAAPNRSYRY 422
 DB 363 VVLAAGLVFTLTDVPELMSAPNSQARSEKAFHDQHPFFRINQVILTAAPNRSYRY 422
 QY 423 DILLGPKKPSGLIDLLLELELEOERLRLHOMVSPKORNTSLDICIYAPLNPNTSL 482
 DB 423 DILLGPKKPSGLISLDLQELLEOERLRLHOMVSHKORNTSLDICIYAPLNPNTSL 482
 QY 483 YDCINSLIQYQNNNTLLLTANQTLMGQTSQVMDKDFLYCANAPLTFKDGATALASC 542
 DB 483 TDCCNVSLIQYQNNNTLLLTANQTLMGQTSQVMDKDFLYCANAPLTFKDGATALASC 542
 QY 543 MADYGAPVFPFLAIGSYKGDYSEAEALIMTSLNNYPAGDPRLAQAKLMEAFLEEMRA 602

DB 543 IADYGAPVFPFLAIGSYKGDYSEAEALIMTSLNNYPADDPKMAHKLMEAFLEEMRA 602
 QY 603 FORNAGMFOVYFTFAERSLDEINRTTAEDLPFPATSYVIFLYISLALGSYSNSRVNV 662
 DB 603 FORNAGMFOVYFTFAERSLDEINRTTAEDLPFPATSYVIFLYISLALGSYSNSRVNV 662
 QY 663 DSKATLGLGVAVVVGAAVMAAMGFPSYSGIRSSLYLVQVPPVLUSVQADNIFIVLEXY 722
 DB 663 DSKATLGLGVAVVVGAAVMAAMGFPSYSGIRSSLYLVQVPPVLUSVQADNIFIVLEXY 722
 QY 723 RLPRRPGSPREVIHGRALGRVAPSWMLGSLSEAIQFGLATPMPAVTTPALTSGLAAYL 782
 DB 723 RLPRRPGSPREVIHGRALGRVAPSWMLGSLSEAIQFGLATPMPAVTTPALTSGLAAYL 782
 QY 783 DFLQMSAFVALLSLDSRKQASRLDVCCVYKQELPPPGQEGILLGFQKAVAPFLH 842
 DB 783 DFLQMSAFVALLSLDSRKQASRLDVCCVYKQELPPPGQEGILLGFQKAVAPFLH 842
 QY 843 WITRGVILLPLAFGVSLYSKCHISVGLDDELAPKDSYLDYFLNRIFEVGAARYF 902
 DB 843 WITRGVILLPLAFGVSLYSKCHISVGLDDELAPKDSYLDYFLNRIFEVGAARYF 902
 QY 903 VTTGLVNSSEAGMNAICSSAGCNNEFTOKIOVATEPPEBQSYALIPASSWVDFIDWLT 962
 DB 903 VTTGLVNSSEAGMNAICSSAGCNNEFTOKIOVATEPPEBQSYALIPASSWVDFIDWLT 962
 QY 963 P-SSCCRLVYISGPNKDKCPTVNSLNCIAKCMISITMGSVRPSVQFHKYLPWFLNDRPN 1021
 DB 963 P-SSCCRLVYISGPNKDKCPTVNSLNCIAKCMISITMGSVRPSVQFHKYLPWFLNDRPN 1021
 QY 1022 IKCPKGLAAVSTSVNLSDQVTLASRFMAVHKPLKNSQDYTBALRAARELANITADLR 1081
 DB 1022 IKCPKGLAAVSTSVNLSDQVTLASRFMAVHKPLKNSQDYTBALRAARELANITADLR 1081
 QY 1023 IRCPKGLAAVSTSVNLSDQVTLASRFMAVHKPLKNSQDYTBALRAARELANITADLR 1082
 DB 1023 IRCPKGLAAVSTSVNLSDQVTLASRFMAVHKPLKNSQDYTBALRAARELANITADLR 1082
 QY 1082 KVPGRDPAPEVPEPYITNVFBOYVTLIPBGLFMLSGLVFPFAVSCLLGLDLSGLLN 1141
 DB 1082 KVPGRDPAPEVPEPYITNVFBOYVTLIPBGLFMLSGLVFPFAVSCLLGLDLSGLLN 1141
 QY 1142 LLSIVMILVDVTFGMALDISNAVSLINLVASGMSVEFYSHIRSPAISTKPTMLERA 1201
 DB 1142 LLSIVMILVDVTFGMALDISNAVSLINLVASGMSVEFYSHIRSPAISTKPTMLERA 1201
 QY 1202 KEATISGSAVPAVAMTNLPGILVLGAKAQLOIIFFRNLNLTLLGLHGLVFLPYT 1261
 DB 1202 KEATISGSAVPAVAMTNLPGILVLGAKAQLOIIFFRNLNLTLLGLHGLVFLPYT 1261
 QY 1263 LSYGPDVNPALALEQKRAEEVAAVMAVASCNPHESTRVSTADNIVYNSPESG-1KGAGA 1320
 DB 1263 LSYGPDVNPALALEQKRAEEVAAVMAVASCNPHESTRVSTADNIVYNSPESG-1KGAGA 1320
 QY 1321 ISNPLPNNGROP 1332
 DB 1321 ISNPLPNNGROP 1332
 QY 1332 ASSSLPKSDQKF 1333
 DB 1332 ASSSLPKSDQKF 1333

RESULT 15
 ADY60748 ID ADY60748 standard; protein; 1333 AA.
 XX ADY60748;
 XX 19-MAY-2005 (first entry)
 XX
 DE Mouse Niemann-Pick C1-like protein 1 (NPC1L1) protein.
 XX
 KW Niemann-Pick C1-like protein 1; Antihypemic; Antiartherosclerotic;
 KW Cardiant; Vasotrophic; Cerebroprotective; cholesterol; hyperlipidemia;
 KW metabolic disorder; atherosclerosis; cardiovascular disease;
 KW coronary artery disease; cerebrovascular ischemia; neurological disease;
 KW arteriosclerosis; transgenic animal.
 XX
 OS Mus sp.

XX WO2005015988-A1.
 PN 24-FEB-2005.
 XX 16-DEC-2003; 2003WO-US040113.
 XX 17-JUL-2003; 2003US-00621758.
 PR 22-AUG-2003; 2003US-00646301.
 PR 16-SEP-2003; 2003US-00663208.
 XX (SCHR) SCHERING CORP.
 PA
 XX
 PI Altmann SW, Murgolo NJ, Wang LQ, Graziano MP;
 XX MPI, 2005-284403/29.
 DR N-PSDB; ADY60747.
 XX
 PT Novel isolated NPC1L1 polypeptide, useful for identifying antagonist of
 PT NPC1L1 that inhibits NPC1L1 mediated sterol or 5 alpha-sterol uptake, for
 PT treating hyperlipidemia or atherosclerosis in mouse, rat or human.
 XX
 PS Claim 30; SEQ ID NO 12; 146pp; English.
 XX
 XX The invention relates to an isolated Niemann-Pick C1-like protein 1
 CC (NPC1L1) polypeptide comprising 42 or more contiguous amino acids from an
 CC amino acid sequence appearing as ADY60738, ADY60740 or ADY60748 (Rat,
 CC human and Mouse NPC1L1). Also included are an isolated polynucleotide
 CC encoding NPC1L1 (appearing as ADY60737, ADY60739 or ADY60747), a
 CC recombinant vector comprising the polynucleotide, a host cell comprising
 CC the vector, an isolated antibody which specifically binds to NPC1L1 (or
 CC to a rat NPC1L1 peptide sequence appearing as ADY60775-ADY60778),
 CC preparation of NPC1L1, a mutant transgenic mouse comprising a homozygous
 CC mutation of endogenous chromosomal NPC1L1 (where the mouse does not
 CC produce any functional NPC1L1 protein), an offspring or progeny of the
 CC mouse having inherited a mutated NPC1L1 allele of the mouse, screening a
 CC sample for an intestinal sterol or 5alpha-sterol absorption antagonist,
 CC inhibiting NPC1L1 mediated sterol or 5alpha-sterol uptake in a subject, a
 CC kit (comprising ezetimibe or substituted azetidinone in a pharmaceutical
 CC dosage form, and information indicating that NPC1L1 is a target of
 CC ezetimibe or substituted azetidinone), decreasing the level of intestinal
 CC sterol or salpha-sterol absorption in a subject (involving reducing the
 CC level of expression of NPC1L1 in the subject), identifying an antagonist
 CC of NPC1L1 and an isolated mammalian cell which lacks a gene encoding a
 CC functional NPC1L1 protein. NPC1L1 is useful for identifying an antagonist
 CC of NPC1L1 which is useful for inhibiting or decreasing the level of
 CC NPC1L1 mediated sterol or salpha-sterol uptake in a subject such as
 CC mouse, rat or human. The level of expression of NPC1L1 in the subject is
 CC reduced by mutating NPC1L1 in the subject. The NPC1L1 knockout mouse is
 CC useful for screening a sample for intestinal sterol or salpha-sterol
 CC absorption antagonist. The NPC1L1 antagonists identified are useful for
 CC treating medical conditions caused or mediated by NPC1L1, e.g.,
 CC hyperlipidemia, atherosclerosis, coronary heart diseases, stroke or
 CC arteriosclerosis. The present sequence represents a mammalian NPC1L1
 CC protein.
 XX
 XX
 SO Sequence 1333 AA;
 Query Match 78.3%; Score 5407; DB 9; Length 1333;
 Best Local Similarity 77.2%; Pred. No. 0;
 Matches 1028; Conservative 126; Mismatches 174; Indels 4; Gaps 4;

QY 183 LAVGTCGVYGSALCNAQRWLNFGDGTGNGLAFLDITPHLLBFGQAVSGIOPLENGVAR 242
 183 LAVGTCGVYGSALCNAQRWLNFGDGTGNGLAFLDITPHLLBFGQAVSGIOPLENGVAR 242
 Db 183 LAVGTCGVYGSALCNAQRWLNFGDGTGNGLAFLDITPHLLBFGQAVSGIOPLENGVAR 242
 QY 243 CNEGQDDVATCGCQCCAAACSCPAIARPOLDSFFYLGOMGSLVLIITLCSAVAVTILL 302
 243 CNEGQDDVATCGCQCCAAACSCPAIARPOLDSFFYLGOMGSLVLIITLCSAVAVTILL 302
 Db 243 CNEGQDDVATCGCQCCAAACSCPAIARPOLDSFFYLGOMGSLVLIITLCSAVAVTILL 302
 QY 303 VGFRAVAPDKSGMVDPKGTSLSDLSTSTHLLLOFGQGMGTWVSWPFLTIVLSVIP 362
 303 VGFRAVAPDKSGMVDPKGTSLSDLSTSTHLLLOFGQGMGTWVSWPFLTIVLSVIP 362
 Db 303 VGFRAVAPDKSGMVDPKGTSLSDLSTSTHLLLOFGQGMGTWVSWPFLTIVLSVIP 362
 QY 363 VVALAGVFTLTTDPVLELMSAPNSQASEKAFFHDQHPFRRTNOVILTAPRRSSRY 422
 363 VVALAGVFTLTTDPVLELMSAPNSQASEKAFFHDQHPFRRTNOVILTAPRRSSRY 422
 Db 363 VVALAGVFTLTTDPVLELMSAPNSQASEKAFFHDQHPFRRTNOVILTAPRRSSRY 422
 QY 423 DSIILGPKNFGSLIDLDLLELLELQBRHLHLOWSEADARNISLQDICTAPLNPNTSL 482
 423 DSIILGPKNFGSLIDLDLLELLELQBRHLHLOWSEADARNISLQDICTAPLNPNTSL 482
 Db 423 DSIILGPKNFGSLIDLDLLELLELQBRHLHLOWSEADARNISLQDICTAPLNPNTSL 482
 QY 483 YDCINSILQYFQNNRTLLLTANQTLMGQTSQVDMKDFLYCANAPLTFKQDTALASC 542
 483 YDCINSILQYFQNNRTLLLTANQTLMGQTSQVDMKDFLYCANAPLTFKQDTALASC 542
 Db 483 YDCINSILQYFQNNRTLLLTANQTLMGQTSQVDMKDFLYCANAPLTFKQDTALASC 542
 QY 543 MADYGA PVFPFLAIGGYKQDYSEALITWPSLANNYPADPRLAQAQKWEAEFLBEMRA 602
 543 MADYGA PVFPFLAIGGYKQDYSEALITWPSLANNYPADPRLAQAQKWEAEFLBEMRA 602
 Db 543 MADYGA PVFPFLAIGGYKQDYSEALITWPSLANNYPADPRLAQAQKWEAEFLBEMRA 602
 QY 603 FQRMAMGFQVTTAERLEDEINRTTAEDLPFATSYIYFIYISLALCYSYSMSVMV 662
 603 FQRMAMGFQVTTAERLEDEINRTTAEDLPFATSYIYFIYISLALCYSYSMSVMV 662
 Db 603 FQRMAMGFQVTTAERLEDEINRTTAEDLPFATSYIYFIYISLALCYSYSMSVMV 662
 QY 663 DSKATTLGLGVAIVLGNVAMAMGFYSYLRSSVLVLOVPLVLSVADNIFLVEYQ 722
 663 DSKATTLGLGVAIVLGNVAMAMGFYSYLRSSVLVLOVPLVLSVADNIFLVEYQ 722
 Db 663 DSKATTLGLGVAIVLGNVAMAMGFYSYLRSSVLVLOVPLVLSVADNIFLVEYQ 722
 QY 723 RLPRRGEPEPEVHIGRALGVAPSMILCSISEAICEFLGALTMPAVRTALTSGLAVIL 782
 723 RLPRRGEPEPEVHIGRALGVAPSMILCSISEAICEFLGALTMPAVRTALTSGLAVIL 782
 Db 723 RLPRRGEPEPEVHIGRALGVAPSMILCSISEAICEFLGALTMPAVRTALTSGLAVIL 782
 QY 783 DFLLOASFAVALLSDSKROEASRLDYCCVQDELPPEQSGSLIGFQKAYAPRLH 842
 783 DFLLOASFAVALLSDSKROEASRLDYCCVQDELPPEQSGSLIGFQKAYAPRLH 842
 Db 783 DFLLOASFAVALLSDSKROEASRLDYCCVQDELPPEQSGSLIGFQKAYAPRLH 842
 QY 843 WTRGVVLLFLALFGVSLYSMCHISVGLDOELALPDQSLDYFLFLNRYFEVGAUYF 902
 843 WTRGVVLLFLALFGVSLYSMCHISVGLDOELALPDQSLDYFLFLNRYFEVGAUYF 902
 Db 843 WTRGVVLLFLALFGVSLYSMCHISVGLDOELALPDQSLDYFLFLNRYFEVGAUYF 902
 QY 903 VTTGYNFSEAGNNAICSSAGCNNSFTQKIQVTEFPQSYLAIPASSWVDFIDMLT 962
 903 VTTGYNFSEAGNNAICSSAGCNNSFTQKIQVTEFPQSYLAIPASSWVDFIDMLT 962
 Db 903 VTTGYNFSEAGNNAICSSAGCNNSFTQKIQVTEFPQSYLAIPASSWVDFIDMLT 962
 QY 963 P-SSCCRLYISGNKQKFCBSTVNSLNCNKSITWGSVRSVBOFHXYLPFLNDRPN 1021
 963 P-SSCCRLYISGNKQKFCBSTVNSLNCNKSITWGSVRSVBOFHXYLPFLNDRPN 1021
 Db 963 P-SSCCRLYISGNKQKFCBSTVNSLNCNKSITWGSVRSVBOFHXYLPFLNDRPN 1021
 QY 1022 IKCPKGLAAYSTVNLTSQVLAERFMAVYHPELKNQSDPYTALPARAELANITADLR 1081
 1022 IKCPKGLAAYSTVNLTSQVLAERFMAVYHPELKNQSDPYTALPARAELANITADLR 1081
 Db 1022 IKCPKGLAAYSTVNLTSQVLAERFMAVYHPELKNQSDPYTALPARAELANITADLR 1081
 QY 1082 KVPQTDPAFVFPPTTNNVFEBOYLTLTLPGLFMLSCLVPTPAVSCLLGLDLRSGLAN 1141
 1082 KVPQTDPAFVFPPTTNNVFEBOYLTLTLPGLFMLSCLVPTPAVSCLLGLDLRSGLAN 1141
 Db 1082 KVPQTDPAFVFPPTTNNVFEBOYLTLTLPGLFMLSCLVPTPAVSCLLGLDLRSGLAN 1141
 QY 1142 LLSIWMILVDTVGFMAIMDISYNAVSLINLSAVAGMSVEFVSHITRSPALSTKPTWLERA 1201
 1142 LLSIWMILVDTVGFMAIMDISYNAVSLINLSAVAGMSVEFVSHITRSPALSTKPTWLERA 1201
 Db 1142 LLSIWMILVDTVGFMAIMDISYNAVSLINLSAVAGMSVEFVSHITRSPALSTKPTWLERA 1201
 QY 1202 KEATISGSAVPAVGAVMTNIPGLIVGLAQAOLIQIFFRFLNLTITLGLHGLVFLPVI 1261
 1202 KEATISGSAVPAVGAVMTNIPGLIVGLAQAOLIQIFFRFLNLTITLGLHGLVFLPVI 1261
 Db 1202 KEATISGSAVPAVGAVMTNIPGLIVGLAQAOLIQIFFRFLNLTITLGLHGLVFLPVI 1261
 QY 1261 KDATIFMGSAVFAVGAVMTNIPGLIVGLAQAOLIQIFFRFLNLTITLGLHGLVFLPVI 1262
 1261 KDATIFMGSAVFAVGAVMTNIPGLIVGLAQAOLIQIFFRFLNLTITLGLHGLVFLPVI 1262
 Db 1261 KDATIFMGSAVFAVGAVMTNIPGLIVGLAQAOLIQIFFRFLNLTITLGLHGLVFLPVI 1262

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using SW model

Run on: March 22, 2006, 22:55:32 ; Search time 52 Seconds
(without alignments)
2117.769 Million cell updates/sec

Title: US-10-736-769-4

Perfect score: 6909
Sequence: 1 MAEAGLRGMLWALLRLRLAQ.....GSIKAGATSNPLPNNRGPF 1332

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/prodata/1/iaa/5_COMB.pep:*
2: /cgn2_6/prodata/1/iaa/6_COMB.pep:*
3: /cgn2_6/prodata/1/iaa/H_COMB.pep:*
4: /cgn2_6/prodata/1/iaa/PCTUS_COMB.pep:*
5: /cgn2_6/prodata/1/iaa/RE_COMB.pep:*
6: /cgn2_6/prodata/1/iaa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2403	34.8	1318	2	US-09-949-016-10152
2	2402.5	34.8	1278	2	US-09-462-136-2
3	2385	34.5	1319	2	US-09-462-136-4
4	1329	19.2	1170	2	US-09-462-136-6
5	1078	15.6	1296	2	US-09-462-136-9
6	651.5	9.4	1447	1	US-08-540-406-19
7	651.5	9.4	1447	2	US-08-656-055-19
8	651.5	9.4	1447	2	US-08-954-668-19
9	651.5	9.4	1447	2	US-09-268-140-5
10	651.5	9.4	1447	2	US-08-918-658-19
11	651.5	9.4	1447	2	US-09-724-631-19
12	651.5	9.4	1447	2	US-08-954-701A-19
13	651.5	9.4	1447	2	US-09-807-007-6
14	651.5	9.4	1447	2	US-09-754-032-19
15	651.5	9.4	1447	2	US-08-916-140-19
16	651.5	9.4	1447	4	PCT-US95-13233-19
17	646.5	9.4	1434	1	US-08-540-406-10
18	646.5	9.4	1434	2	US-08-656-055-10
19	646.5	9.4	1434	2	US-08-954-668-10
20	646.5	9.4	1434	2	US-08-918-658-10
21	646.5	9.4	1434	2	US-09-724-631-10
22	646.5	9.4	1434	2	US-08-954-701A-10
23	646.5	9.4	1434	2	US-09-754-032-10
24	646.5	9.4	1434	2	US-08-916-140-10
25	646.5	9.4	1434	4	PCT-US95-13233-10
26	639.5	9.3	642	2	US-09-248-796A-16161
27	628	9.1	1296	2	US-08-857-636-60

28	584	8.5	1203	2	US-09-293-505-2	Sequence 2, Appl1
29	584	8.5	1203	2	US-09-060-939A-2	Sequence 2, Appl1
30	583.5	8.4	1182	2	US-09-293-505-7	Sequence 7, Appl1
31	583.5	8.4	1182	2	US-09-060-939A-7	Sequence 7, Appl1
32	581.5	8.4	531	2	US-09-248-796A-16160	Sequence 16160, A
33	580	8.4	1203	2	US-09-207-857-2	Sequence 2, Appl1
34	580	8.4	1203	2	US-09-809-280A-2	Sequence 2, Appl1
35	579	8.4	1146	2	US-09-807-007-1	Sequence 1, Appl1
36	526.5	7.6	1311	1	US-08-540-406-4	Sequence 4, Appl1
37	526.5	7.6	1311	2	US-08-656-055-4	Sequence 4, Appl1
38	526.5	7.6	1311	2	US-08-954-668-4	Sequence 4, Appl1
39	526.5	7.6	1311	2	US-08-918-658-4	Sequence 4, Appl1
40	526.5	7.6	1311	2	US-09-724-631-4	Sequence 4, Appl1
41	526.5	7.6	1311	2	US-08-954-701A-4	Sequence 4, Appl1
42	526.5	7.6	1311	2	US-09-754-032-4	Sequence 4, Appl1
43	526.5	7.6	1311	2	US-08-916-140-4	Sequence 4, Appl1
44	526.5	7.6	1311	4	PCT-US95-13233-4	Sequence 4, Appl1
45	474	6.9	1286	2	US-09-268-140-3	Sequence 3, Appl1

ALIGNMENTS

```
RESULT 1
US-09-949-016-10152
; Sequence 10152, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTNER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10152
; LENGTH: 1318
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10152

Query Match          34.8%; Score 2403; DB 2; Length 1318;
Best Local Similarity 39.7%; Pred. No. 3.9e-218;
Matches 525; Conservative 238; Mismatches 451; Indels 110; Gaps 26;

QY      2 AEEAGLRGMLWA-----LLRLAOSPEYTHHQGYCAFYDEG-----KNEBELSG 47
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB      32 ASAAALGHSMTAGLALGILLLLCPAOVFSQ-----SCWVGECIAVADKRYNCEYSG 86
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY      48 SMTLLNSVSLSTNPARKITGDHLILLOKICPLTYPTQAACCAKSLTASISITK 107
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB      32 ASAAALGHSMTAGLALGILLLLCPAOVFSQ-----SCWVGECIAVADKRYNCEYSG 86
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY      87 -----PKPLPKQDYDVOELCPQFFFG-NVSLCCDVROLOTLKNDQLPL 111
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB      108 ALITRCPCASDNFVNLHCHNTCSPNQSLFINVTR-----VAOLGAGQLPAVVAEAYFQHS 163
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY      132 QPLSRPSCRYNLNLFCELTGSPRSQFLNATVATDYDVPVNTKTNVKELOYVGGS 191
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB      164 PAEQSYDSCGRVVPAAATLAVGTMGCVGSALCNARMLNFGCDPTGNGIAPDIT----- 219
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY      192 FANAMTNACRDVAPSSNDKALGLLCKXADA--CNATNMIKVFNFNDNGQAPPTIPVFS 250
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB      220 ---FHLLRGGQANGSIGIOPINBEGVARNCSGGDVATSCGQDCAACCPAARPO----- 270
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

QY      251 DPEVH-----GMEPMNNAATKGCDSVDEVTAPCCGQDCSYCGRPQPPPPAPW 300
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB      271 ---ALDSTFYLGQPGSLVLIILILCSVFAVVTILL-----VGFRAVPARDKS 314
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
```

Db 301 TILGLDAMVIMMITYMAFLVFFGAFVWVCYRKRYFVSEYTPIDISNIAFSV-NASDGG 359
Qy 315 KMWDEKGTSLSDKLSFSTHTLLGQFPQMGTMVASMPLTILVSVI PVALAAGLVTE 374
Db 360 E-----ASCDDPVSAAFEGCLRLFTRMGSFCVRNPPCVIFPFLVFTTACSSGLVFR 412
Qy 375 LTTDPVELMSAPNSQARSEKAFHDQFPPFTNQVILTAPNRSSYRSDLLGPK-NFS 433
Db 413 VTTNPDLMSAPSSQARLEKEYFDQHFQFPFTEQLIRAPLTDKHIYQPYPSGADVPG 472
Qy 434 GILDDLLELELEGERLRLHLOVMSPEAQBNISLQDICAPLPNDTSLYDCCINSLOY 493
Db 473 PPLDQILHQLVLDQLAIEN--ITASYDNETVTLQDICAPLSPVNT--NCTILSVNY 527
Qy 494 FQNNRTLLLTANQTLMGQTSQVMDKHFLYCANAPLTKDGTALALSCMADYGAPVFP 553
Db 528 FQNSHSVLDHKKGDPPF--VYADYHTHFLYCVAPASLNDTSLHDPCLGTFGGVPFW 584
Qy 554 LAIGYKGDYSBAEALIMTFSLNNYPADPRLAOKLWEAFLBEMRAFORRMAGMFGV 613
Db 585 LVLGQYDDQNNYNTATLVITPPVNNYNDTEKLOQAQWKEKFINFVKYKN--PNLTI 641
Qy 614 TTTAERSLEDEINRTTAEDLPFATSYIYIFLYISLALGSYSWSRWMDSKATLGLGV 673
Db 642 SFTAERSLEDEINRESDSVFTVVISYALMFLYISLALGHKSCRLVDSKVSIGIAI 701
Qy 674 AVVLGAVMAAMGFPSYIRSSLVLQVPPFLVLSGADNIFLVEYQRLRRPGEPE 733
Db 702 LIVLSSVACSLGVFSYIGPLTIVIEVIFPLVLANGVNIIFLVOAYQDERLQGETLD 761
Qy 734 VHGRLGAVAPSMILCSLEAICFPLGALTMPAVRTALTSGLAVIDLFLQNSAFVA 793
Db 762 QOLGVLGAVASMFSSFSFETVAFPLGALSVMPAVHTSLFAGLAVFIDFLQITCFVS 821
Qy 794 LILSLSKROEARLDVCCVQPELPPGO--GBGILLGFQKAYAPFLHMTTRGVLL 852
Db 822 LILGLDIPKQEKRLDIFCCVRGADGTSVQASESCIFRFKNSYSLLKMDRPLVIAI 881
Qy 853 FLATFVSLYSMCHISVGLDOELALPKDSVLLDYFLNRYFEVGAFFVTTLGYNFS 912
Db 882 FGVGLSFSIAVANKVDIGLDQSLMDDSYMDFYSIQYLHAGFPVFLVEBGHDYS 941
Qy 913 EAGMNAICSSAGCNNEFSFTOKIQVATEPQSYLAI PASSWDDFLDLP--SSCCRLYI 971
Db 942 SKGQNNVCGMGCCNNDLSLVOQIFNAQLNRYTRIGFAPSMDIDYFDVKPOSSCCRV-- 999
Qy 972 SGPNKDKECPSTVNSLNCANCKMST--MGSVPSVQCFHKYLPWFLNDPNIKCPGGA 1030
Db 1000 -DNITDQFCNAGVVPACVR--CRPLPBEKQRPQGGDFMRFLPMFLSDNPNKCGGHA 1057
Qy 1031 AYSTSVN--LTSQGVLAGRFMAVHKPLKNSODYTEALAARELANITADIRKVGCTP 1088
Db 1058 AVSSAVNILLGCTRGATYFMTHTVLOTSDAFIDALKKARLIASNT--ETMGINGS-- 1114
Qy 1089 AEVEPYTTITNVFYEQYLTILPEGLFMLSCLVTPFVASCLLGLDLSRGLNLSIYWI 1148
Db 1115 AYRVEPYSVFYVEQYLTIIDTIRNLGVSIGALFLVTMILLGCELMSGAVIMCATIANV 1174
Qy 1149 LVDYGFMAINDISYAVASLINLVSAVMSVEFVSHITSFPAISTYPTMLBERAKETIYM 1208
Db 1175 LVNMGGVMMWGLSILVAASLVNLMVMSGSISEFCSHITRAFVSMKGSVEBAEBELAHM 1234
Qy 1209 GSAVFAVAMTNLPGLIVLGLAKAOLQIFPFLNLTLLTGLLHGLVLPYLISVYGD 1268
Db 1235 GSSVSIGITLTKFGGIVLAFPAKSQIFQIFPRMYLAMLGATHGILFLPULSLITGS 1294
Qy 1269 VNPA 1272
Db 1295 VNKA 1298

US-09-462-136-2
/ Sequence 2, Application US/09462136
/ Patent No. 6426198
/ GENERAL INFORMATION:
/ APPLICANT: Carleia, et al.
/ TITLE OF INVENTION: Genes for Niemann-Pick type C Disease
/ FILE REFERENCE: 4239-53894
/ CURRENT APPLICATION NUMBER: US/09/462,136
/ CURRENT FILING DATE: 2000-06-01
/ PRIOR APPLICATION NUMBER: PCT/US98/13862
/ PRIOR FILING DATE: 1998-07-02
/ PRIOR APPLICATION NUMBER: US 60/051,682
/ PRIOR FILING DATE: 1997-07-03
/ NUMBER OF SEQ ID NOS: 13
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 2
/ LENGTH: 1278
/ TYPE: PRN
/ ORGANISM: Homo sapiens
US-09-462-136-2
Query Match 34.8%; Score 2402.5; DB 2; Length 1278;
Best Local Similarity 39.8%; Pred. No. 4.2e-218;
Matches 522; Conservative 237; Mismatches 450; Indels 103; Gaps 25;
Qy 7 RGMILLALLRLAQSEPYTTIHQPGYCAFYDECG-----KNPELGSGLMTLSNVGCLS 59
Db 4 RGLALGLLTLCPAQVFSQ-----SCWYGECCGAYGDKRYNCESG----- 46
Qy 60 NTPAKITGDHILLOKICPRLYGPNQACCSAKQLVLSLEASLITKALLTRCPACSDN 119
Db 47 --PPPLPDGYDVLQVQLCPGFFFG-NVSLCCDVQQLTLDKNDLPLQFISRCPSCFYN 103
Qy 120 FVNLCHNTCSNQSLEFNTVTR---VAQLGAGQLPAVAAVEAFYQHSFAEQSYDSCSRV 175
Db 104 LNLCELTCSRQSGFLNTATBEDYDPTNQTNTVKEIQYVYGQFANMVAACHDV 163
Qy 176 RYPAATTAAGTMCVGSALCNAQRWLNFGQDGTNGIAPLDT-----FHLLPQQA 228
Db 164 EAPSSNDALGLLCSKADA--CNATNMIEMYNKNDGQAPFTITVPFSDFPVH----- 215
Qy 229 VQSGIQLPNEGVARNEQGDVATCSGDCDCAASPAIARP-----ALDSTFVIG 279
Db 216 ---GMEPMNNATKGCDESVDVETAPCSCDSCSYVCGPKPQPPPPAPMTTILGLDMVYIM 272
Qy 280 QWPGSLVLIILICSVFAVTTL-----VGFVAPARDKAMVDPKGTSL 326
Db 273 WTTVMAFLVFFGAFFAVWVCYRKRYFVSEYTPIDISNIAFSV-NASDKE-----ASC 324
Qy 327 DKLSTHTLLGQFPQMGTMVASMPLTILVSVI PVALAAGLVTELTDPVELMSAP 386
Db 325 DVSVAAFEGCLRLFTRMGSFCVRNPPCVIFPFLVFTTACSSGLVFPVTPNPDVMSAP 384
Qy 387 NSQASSEKAFHDQHGPPFTNQVILTAPNRSSYRSDLLGPK-NFSGIILDDLLELL 445
Db 385 SQQARLEKEYFDQHFQFPFTEQLIRAPLTDKHIYQPYPSGADVFPPLDQILHQLV 444
Qy 446 ELQERLRLHLOVMSPEAQBNISLQDICAPLPNDTSLYDCCINSLOYFQNNRTLLTLTA 505
Db 445 DLQAIEN--ITASYDNETVTLQDICAPLSPVNT--NCTILSVNLFQNSHSLVDHKK 499
Qy 506 NOTLMGQTSQVMDKHFLYCANAPLTKDGTALALSCMADYGAPVFPPLAIGYKGDYS 565
Db 500 GDDPF--VYADYHTHFLYCVAPASLNDTSLHDPCLGTFGGVPFWVLGQYDDQYN 556
Qy 566 EAEALIMTFSLNNYPAGPRLAOKLWEAFLBEMRAFORRMAGMFGVTFPTAERIEBEL 625
Db 557 NATALVITPPVNNYNDTEKLOQAQWKEKFINFVKYKN--PNLTIISFTAERSIEBEL 613
Qy 626 NRTTAEDLPFATSYIYIFLYISLALGSYSWSRWMDSKATLGLGVAVVLGAVMAAMG 685
Db 614 NRESDSVFTVVISYALMFLYISLALGHKSCRLVDSKVSIGIAGLILVLSVACSLG 673

Qy	686	PFSTVLGRSSVLIVQVLPVLVSLVSGADNIFLVLVYQRLPRRPREPEVHIGALGVAP	745
Db	674	VFSYVIGLPPLTIVLEIVLPFLVAVGVNIFLVOAYQDRDELQSEITLDQQLGRVLGSVP	733
Qy	746	SMILCSSEALCFPLGALTPEWPAVRTFALVTSGLAVIILDFLLQMAFAVALSLDSKREAS	805
Db	734	SMELSSSEETVAFPLGALSVNPAVHTSFLPAGLVFIDFLQITCFVSLGLDKROEKV	793
Qy	806	RLDVCCCKPPELPPPGQ-GEGLLVGFGEFOKAVAPFLMLHMITRGVLLFLALFSGVLSYSM	864
Db	794	RLDIFCCYRGABDGTSVQASSCLEFRPKNSYSLLLKDMKRPFLVIALFVGVLSFSLAVL	853
Qy	865	CHISVGLDDELALPKDSYLLDFLELNRYFEVGAQVYFVTTLGYNFSSEAGMAICSSAG	924
Db	854	NKVIDIGLDQSLSMEDDGSVMVDYFKFSIQYLHAGPPVYVLEEGHDYTSNGQNMVCGAMG	913
Qy	925	CNNFSFTQKIQYATEPPEQSYLAIIPASSWDDFDIMLTLP-SSCCRLYISGPNKDKCPST	983
Db	914	CNNDSLVOQIIPNAQLDNYTRIGFAPSMSIDYFDWVQPSGCRV--DNITQPCNAS	970
Qy	984	VNSLNLCKNCKMSIT-MGVSRSVQGFHKLYLPMFLNDRPNIKCPRGGLAAVSTVN--LTS	1040
Db	971	VNDPACVR-CRPLTPEGKORQOGDFMFLPMFLSDNPNPKCGGHAAYSSAVNILLGH	1029
Qy	1041	DGQVLASFMAVYHPLKNSODYTEALFRARELANITPADLAKVGTDPAPFVFPYITNV	1100
Db	1030	GTRGATVEMHYHYLVLOTSADPFDLKKARLIASNVT-ETMGINGS--AKRVFPYSFVY	1086
Qy	1101	FYEQYLTITLPEGLFMLSLCLVPFEAVSCLLGLDLRSGLNLISIVMLVDTVGFMAIWD	1160
Db	1087	FYEQYLTITIDTITNLGVLGCAIFLYVMVLLGCELMASVIMCAITAVIIVMFGVMLMG	1146
Qy	1161	ISYNAVSLINIVSAVGVMSVEFVSHITRSFAISTKPYMLERAKKATISMGSAVPAVAMTN	1220
Db	1147	ISLNAVSLINIVMSGGISVEFCSHITRAFTVSMKGSRYERAEALAHGSSVSFGITLTK	1206
Qy	1221	LPGLVLGLARQLQIIFFRMLNLTITLGLHGLVFLPYLTVSGVDVNA	1272
Db	1207	FGGIVVLAFAKSOIFQIIFYFRMYLAMVLLGATGHLIFLQVLLSYIGPSVNKA	1258
RESULT 3			
US-09-462-136-4			
Sequence 4, Application US/09462136			
Patent No. 6426198			
GENERAL INFORMATION:			
APPLICANT: Carstee, et al.			
TITLE OF INVENTION: Genes for Niemann-Pick Type C Disease			
FILE REFERENCE: 4239-53894			
CURRENT APPLICATION NUMBER: US/09/462,136			
CURRENT FILING DATE: 2000-06-01			
PRIOR APPLICATION NUMBER: PCT/US98/13862			
PRIOR FILING DATE: 1998-07-02			
PRIOR APPLICATION NUMBER: US 60/051,682			
PRIOR FILING DATE: 1997-07-03			
NUMBER OF SEQ ID NOS: 13			
SOFTWARE: Patentin Ver. 2.0			
SEQ ID NO 4			
LENGTH: 1319			
TYPE: PRT			
ORGANISM: Mus sp.			
US-09-462-136-4			
Query Match			
Best Local Similarity 34.5%; Score 2385; DB 2; Length 1319;			
Matches 524; Conservative 231; Mismatches 440; Indels 116; Gaps 28;			
Qy	14	LLRLAQSPPYTTIHQPGYCAFYDECGKNPBLSGSLMTLSNVSCLSNTPAKRTIGDHLIL	73
Db	53	LLLLLCPAQVFSQ-----SCWYGEBCG---IATGD---KRYNCKYSGPPKPLPXDGYDL	100
Qy	74	LQIKCPRLVYTGNTQACCSAQVLVSLASLSITKALLILRCPACGDNFVNLHCHNTCSPNQ	133

Db	101	VQELCPBLFF--DNVSLCCDDIQQLQTLKSNLOPLQFLSRPCSPCYNMLTFCELTCSPHQ	159
Qy	134	SLPFINVRVAGOLGAGOLPA-----VVAEAFYQHSFAQSYDSCGRVFRPAATLAVGTMC	189
Db	160	SQFLNATVATEDYDPKTRPKNTWKLELYVYGAFAMAMTNACRDEBAPSENEKALGLIC	219
Qy	190	GVTGSAICNAQRWLNFGQDTGNGLAPLDI-----THLLEFGQAVSGCIGPLNEGVARN	244
Db	220	GRDARA--CNATNMWLEVMFNKDNQGAFFPTIIPVESDISIL-----GMEPMRNATKCN	270
Qy	245	ESQGDVATGSCQDCAASCALARPQMLDSTFYLQMGPSGLVLIILCSVPFVAVTLLVG	304
Db	271	ESVDEVTPGSCDCDSIVCG--PRPOLLQIGGGGGM--GLDAMVIMWTVYAFLEFFVG	327
Qy	305	FRVA-----PARDKSNWVDPKKGTSLSDKTSFSTHTLLGOF	340
Db	328	ALLAVWCHRRRYFVSEYTPIDSNIAFSVNSDKE-----ASCCDPLGAFPDCLIRM	380
Qy	341	FOGNGTWASWPLTLLVLSVPVVALAAGVFTLETTDPVELMSAPNSQARSEKAFDHOH	400
Db	381	FTTKGACVARNPTCIIFFSLAFITVCGSGLVQVYOTTNPPELMSAPHSQARLEKEYYDKH	440
Qy	401	FGPFRTNOVLTLPANRNSRYKXDSLIGPK--NFSGLIDDLLELLELQERLRLQVWSP	459
Db	441	FGPFRTREQLIQAPNNSVHIYEYEPAGADVFPQPLNKEILHQVNLIQ-----IAIESI	495
Qy	460	EAQRN---ISLQDTCVAPLNPDTSLYDCCINSLLQYFQNNRLLTLTANOTLMGQTSQV	516
Db	496	TASYNNEVTVLQDICVAPLSPPYNK--NCTIMSVLANFOUSHAVL-----DSQV	541
Qy	517	-----DKMDHLYCANAPLPTFKDQDTALALSGMDLYGAPVPEFLAIGYKKGDEBAE	568
Db	542	GDDEYIYADVHTHPLVCVAPASLNDTSLHGPCLTGGFGEVFPWLVAGDDQNNYNTAT	601
Qy	569	ALINTFSLNNYPAGDPLPAOKLMEBAFLEBMAFQORMAGMFOVTFEAERSLDEINRT	628
Db	602	ALVTTFPVNNYNDTERLQRAMAMEKEFISFVKYKA---PNLTISTAERSIDELNRE	658
Qy	629	TAEDLPITFATSYIYIFLYISLALGYSWSRWVDSKATLGLGSAVVLGAVMAMAGFS	688
Db	659	SNSDVFYIISYVMPFLYISLALGHIOGCRRLVDKSIKIGAINILLVSSVACSLGIFS	718
Qy	689	YLGRSSVLILQVVPFLVLSVGNINIFIPLYEYQRLRRRQPEPREVHIGRALGVADPSML	748
Db	719	YMGPRLLTILEVIPLEVLAVGVDNITFLVQTYGRDBRLQEBETLDQGLRILGEVAPTMF	778
Qy	749	LCSISEAICEFLGALTMPAVRTFATLSGLAVIIDLQMSAFYALLSLDSKROEASRLD	808
Db	779	LSPSETSAPFFGALSSMPAVHTFSLPAGMAVILIDFLQITCFVSLGLDIIKROEKHLD	838
Qy	809	VCCVCYKQOELRPPQOQ---EGLLLGFQKAYAPFLHLMTITRGVULLFLALFQVLSYM	864
Db	839	ILCCVRGAD---DQCGHASESYLFRFFKVPFAPLLKMDLRPIVAVFVGVLSSFVAVV	895
Qy	865	CHISVGIDQELALPKDSYLLDYFLFLNRYPEVGAQVYFVTTLIGNFSSBAGMANICSSAG	924
Db	896	NKVGIDGQDSISMNDSTVIANFSLAQYLHSGPPVFLBEGYNYSKRCQNVCCGGMG	955
Qy	925	CNNFSFQOKLOYATEPEBQSYALAPASWVDFIDWLTP--SSCCRLYISGPNKQKFCPST	983
Db	956	CDNDSLYOQIINAAELDTYTRVGRAPBSMWIDYDWSPOSSCCRLY---NVTHQFCNAS	1012
Qy	984	VNSINCLKNCKSIT--MGSVRSPEVQFHKXYLPWFLNDRPNIKCPRGGLAIVSTVNLTSQG	1042
Db	1013	VMDPLCYR--GRPLRPEGKQRPQGEKFMKFLPMFLSDNPNPCKCGKGHAAVSAVNIQDD	1071
Qy	1043	QVL--ASRPMAYHKLXKSQDYTELRAARELANIITADLRVPTGDPAFVFPPTITNVF	1101
Db	1072	TYIGGYEMTHTTIKTSADYTDAMKCARLTIASNITEITMRS--KQSD--YRVFYSVFPYVF	1128
Qy	1102	YEOVTLTIPBGLPMLSLCTVPTFVASCLLGLDRSGLLNLSTVMLLVDVVGMLAMD	1161
Db	1129	YEOVTLTIDDTIFMLSVSLGSIPLVTLLVAGCELSMAVIMCTITAMLLVNNFGVMMIIGI	1188

Query Match	15.6%;	Score 1078;	DB 2;	Length 1296;
Best Local Similarity	24.0%;	Pred. No. 2.8e-92;		
Matches 321;	Conservative 237;	Mismatches 492;	Indels 288;	Gaps 41;

QY	KICPELYTGPNTQACCSAKOIVSYLPAISITLALTALTRCPACSDNFWNLCHNTGSPNSL	135
Db	63 EFCPHLLTGDN-KLCTSPSQABGLTKQJLAQARHILGRKPSCTDNRKAKIMCETGSPNOD	121
QY	136 FINVTRVAQL--GAGQAPVAVVAEAF-----YQHS--PAEOSYDSCSRVPAATLAVG	186
Db	122 FVSISEMPRIEKKEGFTPEYOPAEAYVTVTEYRLSTDPAEGFSSCKDVTFGQPALRV-	180
QY	187 TMCVYGGALCNAGWMLNFGDGTNGL-APLDITHLLEPGQAVSGIQP-LINEVACN	244
Db	181 -MC--TSTPCLTNWLEFTGYQNDLNLPIHTKFLYDPIKTPPSDRSTYNNVNFQCD	236
QY	245 ESQGDVAVTSCCODCAACPAIARFQALDSTFYLCOM-----PGSLVHIIILCSYFAVVT	299
Db	237 KSARQWMACTSTSEC--NKEEYANLIDDDGKTSQTCNQHIALCNLFVMLAIFGLSLAV	294
QY	300 ILLVGF-----RVAPRDKSKVNDPKGTSLSDKLSFSTHLLGQF	341
Db	295 LLICGFVFTSYDEDTYNLRQOTSGEESPURNIK-----RTGAMI-----HNMENTNA	342
QY	342 QGQWTVASWPLTILVLSVIPVVALAGIVFELTDPPELMSAPNSQARSEKAPHDOHF	401
Db	343 KDIGMAGRNPKSHFFTCGAVLIFCLPGHITKKESTNVADMMSSRSRSPRQOBEVFNANF	402
QY	402 GPFFRTNOVILIAPRRSSRYSDSLILGPKNFEGIILDLLELLELOERLRLQVSSPEA	461
Db	403 GRPQRYQOIMLL--SHRDPQSGKLYG-----VHKQIIFELPILNAIKNISQDSBG	455
QY	462 QRNISLQDI CIAPLNDPNTSLDYCCINSILQYFQNNRTLLILITANQT-----L	509
Db	456 -RTILDDVVCYKPMRPG-----YDCLMSPTNYFOQNKHEHLDMKSKNEETVEDDAPYF	510
QY	510 MGQTSQVQWMOHFLVCANAPLTFKFKGTALALSCMDYGAVPVPRPLAIGYKQKOVSEEA	569
Db	511 SSEATTDMMNNMAACIDQPMNQK--TYSGLSCWGTGYGSPSAPNN--VQEKNSTYNQALNS	567
QY	570 LIMFSLNNYPAGDEPLLAQAKLMBEAFLEEMAFORMAGMFQVTFTEBRSLEDEINRTT	629
Db	568 IMNTILVYQ--RTEBEIQAKELMEKEFLKCEYREKSPKVI--FSFMAERSTIDIEINDA	624
QY	630 ABDELPIFATSYIVILYISLALGSY-----SSWSRVWDKSKATLGLGQVAVVLGAVMAM	684
Db	625 KDEITVVAIALAFLIGYVTFYSIGRFVFCENQJMS--ILVHS-----	664
QY	685 GPFYSYIGRSSLVILQVVPFLVLSVGADNIFFLVLEYQRLPRRQEPREVNIGRALGRVA	744
Db	665 GF-----	666
QY	745 PSMLLCSLSEALCFPLGALTMPMAVRTFALTSGLAVIDLFDLQMSAFVALLSLDSKROEA	804
Db	667 -----TDLPAIRTFCLVYAGLAVLIDVNLHCTIFLALFWMDQRELN	707
QY	805 SRLDVCCCKPEBELPPQGGEGU-----LLG-----FQCAIYAPFLHMTTRGV	849
Db	708 G-----KP-EFFPFYQIKDLGAYILIGRQATDTEWTOFHNQOVAPFLHMRHRTIIT	758
QY	850 LLLFLALFGVSLYXSMCHISVGLDOELAPKQSYLNDYFLFLNRYFEVGAVPYFVTTLGN	909
Db	759 GIIFLASFTTYIILSKISVGFQDSMAFPEKSYISTHRYLXDKFDVQGPVVFYVYDGBELD	818
QY	910 FSSSEAGMNAICSSAGCNNSFTQKIQYATEPEQSYLALPASSWVDFIDWLTF-DSSCCR	968
Db	819 WHPRPVQNKFTCFPGCSDTSPFNINMVAAGHTEQYTLGSEMYNMWIDYLEMTISRSKPCCK	878
QY	969 LYSISPNDKFEPSTVNSL-----NCLKNQMSITMGSVSPSVQGHKYLPMFLNDRPNIKC	1022
Db	879 VYVHOPN--TFQSTNRNKSALDDKACRTCMDP-----	908

```

QY 1025 KKGJLAAYSTSVNLTPSGOVLAAEFMYAHPYPL--KNSODYTEALRAAREJLAANITADLRK 1082
Db 909 --DRAAPKODIISTFSRGRIOAOSPFMHFKKLISNSSDPIKADTDKRWMSRRLERST-- 964
QY 1083 VPTGDPAAFEVEPPYITITWVEEQYUYYTILPEGLFMJSLCLVPTFAVSCILLGLDRSGJLNL 1142
Db 965 ----DTRAHFAYASKITPPFEQOYSTIMPILTQLOFIVVGVEFGILCYTLGIDVKGAACAV 1020
QY 1143 LSIYMIILVDTVGFMAJWDISYNASVLINLVASGMSVEFWSHITRSPALISTKPTWLERAK 1202
Db 1021 ICQYS-----NYPHIVSS-GILIEFSVNLKGVACSLRQAKORAE 1060
QY 1203 EATISMGSAVFAVAMTNLPEGLIVLGLAKKQLOIPEFFRLNLTLTGLLGLVFLPEVIL 1262
Db 1061 STVSGISGIIILSGPVVITMAGSTWMLPSGAHQLIIVVEFFKFLFLITIVSSAVHALITLIPLL 1120
QY 1263 SYVGPDPVNPALALEQKRAAEVAALVWMAVMAVSCNHSRSTADNIYVN-----HSPEGS 1314
Db 1121 AFGSGRGGSESTSTNDNDEQHDACVLS--PTAESHLSNVEEGILNRPDLIDASHIIDLPL 1178
QY 1315 IKGAGAISNFLPNNGROF 1332
Db 1179 LKAEGGIDKAI--GRDF 1193

```

```

US-08-540-406-19
/ RESULT 6
/ Sequence 19, Application US/08540406
/ Patent No. 5837538
/ GENERAL INFORMATION:
/ APPLICANT: SCOTT, MATHEW P
/ APPLICANT: GOODRICH, LISA V
/ APPLICANT: JOHNSON, RONALD L
/ TITLE OF INVENTION: Patched Genes and their Use
/ NUMBER OF SEQUENCES: 19
/ CORRESPONDENCE ADDRESS:
/ ADDRESSSE: Flehr, Hohbach, Test, Albritton & Herbert
/ STREET: Four Embarcadero Center, Suite 3400
/ CITY: San Francisco
/ STATE: CA
/ COUNTRY: US
/ ZIP: 94111
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patentin Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/540,406
/ FILING DATE: 06-OCT-1995
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Rowland, Bertram I
/ REGISTRATION NUMBER: 20015
/ REFERENCE/DOCKET NUMBER: a60190-1
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 415-781-1989
/ TELEFAX: 415-398-3249
/ INFORMATION FOR SEQ ID NO: 19:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1447 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
US-08-540-406-19

```

Query Match 9.4% Score 651.5; DB 1; Length 1447;
Beet Local Similarity 22.0%; Pred.No.1.1e-51;
Matches 300; Conservative 181; Mismatches 460; Indels 421; Gaps 47;

QY 225 PGAGVSGISGPIANGVGARCNESGGDDVATCCSCDCAASCAPAIARPQALDSFFLYGGMPS 284
|||::|||:: |||:: |||::

```

Db      24 PCRPAAGGRRRTGCGLRRAAARD-----YLRBSYCDAAFALEOI----65
Qy      285 LVLIILLCSVFAVVITLLVGFRAVAPARDKSKWVDPK-----KGTSLSDKLSFSTHTLL 337
Db      66 -----SKGKATGRKAPLMRLAKFKGRLFLKGLCYIQKNC 98
Qy      338 GGFPGCGWGTWVAWPLTITLVSVTPVVALAAGLVFELTDDPVELMSAPNSQARSKAH 397
Db      99 GRF-----LVVGLLIFGAFVAGLKAANLETNVELWVEVGRVSRRLANT 143
Qy      398 DOHFQGFPRTN-QVILTAPNRSSYRDSLLGKNSGILLDILLLELE--IQEHLRLH 454
Db      144 RQKIGBEAMFNPOLMTQTPKEBS-----ANVLTTEALQHLDSALQASRAV 190
Qy      455 QVWSPBAONISLQDICYAP-----LNPNTSLYDCINSLSLQVFQNNRTLLTLTA 505
Db      191 YWV-----RQWKLHCHCYKSGELITETGVMDQILEVLPCLITPLDCFWEGAKLQSGTA 246
Qy      506 NOTLMGQ-----TGQVMDKHPLY-----CAN-----527
Db      247 --YLLGKPLRWTFNPDPLFLEBELKKINYQVDSWEMLNKAEGVGYMDRPCILNPADPDC 304
Qy      528 -----APL-----TFKQGT-----ALALSCMA 544
Db      305 PATAPNKASTKPLDUALVTLNGCGHLSRKYMKQEBELVIGTVKNSGTGLVSHALQTNF 364
Qy      545 DYGAVPFPLAIGGYKDKDYSEAEALIMTFSLNNYPADGPRLAQAALMEAPLEEMRAFO 604
Db      365 QLMTPKQWY--EHFQGYEY-----VSHNNME--DKAAAILLEAWQRTYEVVHQSV 411
Qy      605 RRMAGHFQVTFPAERLEDEINFTTAEPLPIFATSTIVFLYISLALSGYSSWSRWVDS 664
Db      412 AONSTOKVLSFTT-TLLDILKSPSDSVYIRVASGYLMLAAVACLTM--LRMD--CSRS 465
Qy      665 KATLGLGVAVVYGAAMAMGPFYSYLSLITLVNPEPLVSGADNIFIVLEQRL 724
Db      466 QGAVGLAGLVLAALVAAGIGLCSLIGISPNAATTOVLPPLALGCVDDVDFLAHAASGT 525
Qy      725 PRPGEPRVHIGRALGVAPSMILCSLSEALCFPLGALTMPMAVTFALTSGLAVILDF 784
Db      526 GQNKRIPEDRDGECLKTGASVALGISINVTAFMAAILPIPALAFAFLQAAYVVVNF 585
Qy      785 LLOMSAFVALLSLDSKROBASRLDVCC-----VKROEL-----PPRG 822
Db      586 AAVLLIFPALISMDLYRREDRLDICFTSPCVSRVIOVEPOAYTDTHTNTRYSPBPY 645
Qy      823 QGEGL-----827
Db      646 SSHSFAHETQIMQSTIVQLRTEYDPTHVYTTAERSEISVQPVTVTQDTLSCOSPST 705
Qy      828 -----LGGFQKAVAPFLLMHTIRGVVLLPLALFGVSL 861
Db      706 SSTRDLLSQFSDSLHCLBPCCKWTLSSPAEHGVAPFLKKPRAKVIVIFLGLGVL 765
Qy      862 YSKCHISVGLDOELAPKDSYLLDYFLFLNRYFEVGAQVYFTYTLGYNFSSBAQNAIICS 921
Db      766 YGTRVRDLDLDLDIVRRETRYDIFYAOFKXFSS-----YNN-----803
Qy      922 SAGCNNSFTOKIOYAT-----EPFEGSYLAIPASS-----WVDDFDIML-----961
Db      804 -----YIVYQKADYFNIGHLYDLHRSFNKTYLAEKQOLPKKMLHYFRDMLGLGLD 857
Qy      962 -----TPSS-----CCRLYISGPNKDKFCPTVNSLNCIKNCMSITMGS 1000
Db      858 AFDSDMETGKIMPNNNYKNGSDGCVLAKLVGCGSDK--PIDISGLT--KQRLVDADI 913
Qy      1001 VRPSVEQPHKYL-PWFLND-----RPNIKCPKGGLAAY--STSVNLTSDQ 1043
Db      914 INPSA--FYIYLTAWNSNDPVVAAQANIRPHRPFWVHD--ADYMBETRLRIPAAEP 968
Qy      1044 VLASRMAHYKPLKJNSQDYTEALRARABELAANITA-DLAKVPCTDAFEVFPYITINVY 1102
Db      969 IEYAQPFYLANGRLDTSDFEVALEKVRTICSNYTSGLSSYPNG-----YPP--LFW 1018

```

```

Qy      1103 EGYLTLEBGLFMLSCLVPTFAVSCLLGLDLSRGLNLNLSIWMILVDTVGFMAIWDIS 1162
Db      1019 EGYIGLRHMLLFISIVLACFLVCAVFLNMPWTAGII-VWVALMTVELFGMGLIGIK 1077
Qy      1163 YNAVSLINLVASVNGSVFVSHITSF--ALSTRPTWLERAKETISGASVAFVANTN 1220
Db      1078 ISAVPVVILIASVGVEFTVVAALFLTAIGDN--RRAVLALEHMFAPVLDG-AVST 1133
Qy      1221 LFGILVLAGAKQLQIFFPRLNLITLLGLHGLVFLPVLSYVG--PDVNPALALBO- 1277
Db      1134 LIGVLMLAGSEFDFVRYFFAVLALITLGLVNLGLVLLPVLLSFPGFPFVPSFANGRL 1193
Qy      1278 --KRAEVAVAAMVASCPNBPGRVSTADNIYVNSFECSIKG 1317
Db      1194 PTPSPPEPPSVVRFAMPQHTH--SGSDSDSEYSSQTTVSG 1233

```

RESULT 7

US-08-656-055-19

; Sequence 19, Application US/08656055

; Patent No. 6027882

; GENERAL INFORMATION:

; APPLICANT: SCOTT, MATHEW P

; APPLICANT: GOODRICH, LISA V

; APPLICANT: JOHNSON, RONALD L

; TITLE OF INVENTION: Patched Genes and their Use

; NUMBER OF SEQUENCES: 19

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert

; STREET: Four Embarcadero Center, Suite 3400

; CITY: San Francisco

; STATE: CA

; COUNTRY: US

; ZIP: 94111

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/656,055

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/540,406

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Rowland, Bertlem I

; REGISTRATION NUMBER: 20015

; REFERENCE/DOCKET NUMBER: A60190-1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-781-1989

; TELEFAX: 415-398-3249

; INFORMATION FOR SEQ ID NO: 19:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1447 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-656-055-19

; Query Match

; Best Local Similarity 22.0%; Score 651.5; DB 2; Length 1447;

; Matches 300; Conservative 181; Mismatches 460; Indels 421; Gaps 47;

```

Qy      225 PQAVGSGIQLPNEGVAECNBSQGDVATCSQDCAACPAIARQALDSTFYLGQMPGS 284
Db      24 PCRPAAGGRRRTGCLRRAAARD-----YLRBSYCDAAFALEOI----65
Qy      285 LVLIILLCSVFAVVITLLVGFRAVAPARDKSKWVDPK-----KGTSLSDKLSFSTHTLL 337

```



```
Db      144 ROKIGBEAMFNQOLMOTPEKEG-----ANVLTTTEALLQHLDSALQASRVHV 190
Qy      455 QVMSPEAOBNISLQDICTAP-----LNPNTSLYDCCINSLLQYFQNNRTLLLLTA 505
Db      191 VMYN-----RQWLEHLCYKSGELITETGYMDQILEVLPCLITPDLDFWEGAKLQSGTA 246
Qy      506 NOTLMGO-----TSQVDMKDFLY-----CAN----- 527
Db      247 --YLLGKPLRWNTNPDLEFLBELKKINYQVDSWEMLNKAEGVGYMDRPCILNADPDC 304
Qy      528 -----APL-----TFKQGT-----ALALSCMA 544
Db      305 PATAPNKSTKPLDMLALVINGGCHLSRKRYMHOEELIVGTVKNSGTGLVSHALQTMF 364
Qy      545 DYGAVPFPPLAIGYKGDYSEAEALIMTFSLNNYPADPRLAQAOKLMEAEFLEEMRAFO 604
Db      365 QLMTPKQWY---EHRFGYEV-----VSHINWNE---DKAAALILEAMQRTYVEVHOSV 411
Qy      605 RRMAGMFOYTFAERSLEDEINRTAEDLPFATSYIVIFLYISLALSGYSMSRVWVDS 664
Db      412 AONSTQKVLSTFT--TTLDILKSPSDVSVIRVAGSYLLMALVACLTM---LRMD--CSKS 465
Qy      665 KATLIGGAVVVLGAAMAMGPFSSYLIGRSSLVIIQVPEVLVSAGADNIFIVLEYORL 724
Db      466 QGAVGLAGVLLVLAAGLGLCSLIGISFNAAITTVLPPLAGVGDVDFVFLAAHAFST 525
Qy      725 PRPGEPREVIHGRALGRVAPSMILCSLSEALCFPLGALTTPMAVTFALTSGAVILDF 784
Db      526 GQNKRIPEFDRGECIKRGASVALISINVTAFPMALILIPALRAFALQAAVVVVFNF 585
Qy      785 LIQMSAFVALLSDSKROBASRLDVCC-----VKPOEL-----PPPG 822
Db      586 ANVLIFPAILSDMLYRREDRLRIDPCFCTSPCSVSHIQVBPQAYTDTHTNTRYSPBPV 645
Qy      823 QGEBGL----- 827
Db      646 SSHSFAHETQIMQSTVOLRTXYDPHTHYVYTTAERSBSISVQPVTTQDTLSCGSPST 705
Qy      828 -----LIGFQKAYAPFLHMTTRGVVLLFLALFGVSL 861
Db      706 SSTROLSQPSDSSLHCLEPCTKMTLSSFAEKHYAPFLPKPKAKVYVFLFLGLGLVSL 765
Qy      862 YSMCHISVGLDOELALPKDSYLLDYFLFLNNRYFEVGAPVYFVTTLGYNFSSEAGNNAICS 921
Db      766 YGTTVRDGLDLTDIVPRETREVDFIAQFKYFSF-----YNN----- 803
Qy      922 SAGCNPFSTQKIQVAT-----EFPEQSYLAIPASS-----WVDFETDML----- 961
Db      804 -----YIYTKADYENIQHLLYDLHRSFSNVKYVMLEENKOLPKXWMLHYFRDMLQGLQD 857
Qy      962 -----TPSS-----CCRLYISGPNKDKFCPTVNSLNLCKNCKMSITWGS 1000
Db      858 AFDSDMENGKIMPNVKNKSGDDCVLAKYLLVQSGSDK--PIDISQLT--KQRLVYADSI 913
Qy      1001 VAPSYEQFHXYL--PMFLND-----RPNIKCPKGGLAAY--STSVNLTSDQ 1043
Db      914 IMPSA--FYIYLTAWVSNDPVAAVSAQANIRPHRPEMVHDK--ADYMBETRLRIPAAEP 968
Qy      1044 VLASRPMAHKPKLKSQDYTEALBARABELAANITA--DLAKVPCTDAFEVFPYTTINNVY 1102
Db      969 IEVAGPFPYLANGRLDTSDEALEKVRTTCSNTSTGLSSYPNG-----YPP---LFW 1018
Qy      1103 EGYLTLPFEGLFMLSLCLVPTFAVSCLLGLDLRSLGMLNLSTVMILVTVGVFMALMDS 1162
Db      1019 EGYIGRHHLLLFISVVLACTPLVCYAVPFLNPTAII--VMYALAMTVELFGMGLIGIK 1077
Qy      1163 YNAVSLINLVSAGVSVSEFVSHITRSF--AISTKPTMLERAKEATTISMSAVFAGVAMTN 1220
Db      1078 LSAVPVILIASGVIGVEFTVHVALAFLTAIGDKN--BRAVALBLEHMPAPVLDG--AVST 1133
Qy      1221 LGGILVTLGAKAQLIOIFFPRLNLTLTGLLHGLVFLPVILSYVG--PDVNPALALBQ- 1277
```

```
Db      1134 LIGVLMAGSEFDFIVRFPAVLAITLTLGVNLGVLLPVLLSFGCPYEPVSPANGLRL 1193
Qy      1278 --KRAEAAVAAVMVASCNHPBSRVSTADNIYVNSFECSIKG 1317
Db      1194 PTPSPPEPPSVYRFAMPQGHTH--SGSDSDSEVSSQTTVSG 1233

RESULT 9
US-09-268-140-5
; Sequence 5, Application US/09268140
; Patent No. 6268176
; GENERAL INFORMATION:
; APPLICANT: Gemmili, Robert M.
; APPLICANT: Dradkin, Harry A.
; TITLE OF INVENTION: TRC8, A GENE RELATED TO THE HEDGEHOG RECEPTOR, PATCHED
; FILE REFERENCE: 93445-00004
; CURRENT APPLICATION NUMBER: US/09/268,140
; PRIOR FILING DATE: 2000-03-12
; PRIOR APPLICATION NUMBER: US 60/077,723
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1447
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-268-140-5

Query Match      9.4%; Score 651.5; DB 2; Length 1447;
Best Local Similarity 22.0%; Pred. No. 1,1e-51;
Matches 300; Conservative 181; Mismatches 460; Indels 421; Gaps 47;

Qy      225 PQGAVGSGIQPLNEGVARCNBSQGDVATCSCQDCMAAPALARPQALDSTFYLGQMPGS 284
Db      24 PGRPAGGGRRRRTGRLRBAAPDRD-----YLHRPSYCDAAFALEQI--- 65
Qy      285 LVLIILISVPAVYTIILVGRVAPAROKSKVNDK-----KTSLSDLKSFSTHLL 337
Db      66 -----SKGATGRKAPLMRAKFORLLFKLCYCIQKNC 98
Qy      338 GQFQGWGTWVSWPDLTILVSVIPVVALAAGLVFTETTPDVELMSAPNSQARSEKAFH 397
Db      99 GKF-----LVGGLIFGAFAVGLKAALETNVEELMVEGVRSBELVYT 143
Qy      398 DQHFPPFRTN-QVILTPANSSYRYDSLGLPKNPSGILDLILLLELLE--LQERLRLH 454
Db      144 RQKIGBEAMFNQOLMOTPEKEG-----ANVLTTTEALLQHLDSALQASRVHV 190
Qy      455 QVMSPEAOBNISLQDICTAP-----LNPNTSLYDCCINSLLQYFQNNRTLLLLTA 505
Db      191 VMYN-----RQWLEHLCYKSGELITETGYMDQILEVLPCLITPDLDFWEGAKLQSGTA 246
Qy      506 NOTLMGO-----TSQVDMKDFLY-----CAN----- 527
Db      247 --YLLGKPLRWNTNPDLEFLBELKKINYQVDSWEMLNKAEGVGYMDRPCILNADPDC 304
Qy      528 -----APL-----TFKQGT-----ALALSCMA 544
Db      305 PATAPNKSTKPLDMLALVINGGCHLSRKRYMHOEELIVGTVKNSGTGLVSHALQTMF 364
Qy      545 DYGAVPFPPLAIGYKGDYSEAEALIMTFSLNNYPADPRLAQAOKLMEAEFLEEMRAFO 604
Db      365 QLMTPKQWY---EHRFGYEV-----VSHINWNE---DKAAALILEAMQRTYVEVHOSV 411
Qy      605 RRMAGMFOYTFAERSLEDEINRTAEDLPFATSYIVIFLYISLALSGYSMSRVWVDS 664
Db      412 AONSTQKVLSTFT--TTLDILKSPSDVSVIRVAGSYLLMALVACLTM---LRMD--CSKS 465
Qy      665 KATLIGGAVVVLGAAMAMGPFSSYLIGRSSLVIIQVPEVLVSAGADNIFIVLEYORL 724
Db      466 QGAVGLAGVLLVLAAGLGLCSLIGISFNAAITTVLPPLAGVGDVDFVFLAAHAFST 525
Qy      725 PRPGEPREVIHGRALGRVAPSMILCSLSEALCFPLGALTTPMAVTFALTSGAVILDF 784
```


Db 526 GGNKRIPEEDRTGECLKTGASVALTISINVTAFPMALIPBALRAESLOAAVVVNF 585
Qy 785 LLQMSAFVALLSLDSKQASRLDVCC-----VKQDEL-----PPRG 822
Db 586 ANVLIFPAIISMDLYRREDRLDIFCCTSPCVSRVIVQVEPQAYDTHDNTRYSPPPY 645
Qy 823 QGEBL----- 827
Db 646 SSHAHEQTQIMQSTVQLRTEVPHTHVYTTAERSEISVQPVTVTDTLSCSPST 705
Qy 828 -----LLGFQKAYAPFLLHMTTRGVLLLFALLEGVL 861
Db 706 SSTRLDLSQSDSLHCLPEPCRTKTLSSFAEKHYAPFLKPKAKVYVFLGLGLVSL 765
Qy 862 YSMCHISVGLDQELAPKDSYLLDYFLPNRYFEVQAPYFVTLGYNSSSEAGNNAICS 921
Db 766 YGTRVRDGLDLTDIYPRFTRDYDFAAQPKYFSF-----YNN----- 803
Qy 922 SAGCNNEFTQKIQYAT-----EPPEQSYLAIPASS-----WVDFIDWL----- 961
Db 804 -----YITOKADYFNIGHLLYDLHRSFSNVYVWLEENKQLPKWMLHYFRDMLQGLQD 857
Qy 962 -----TPSS-----CCRLYISGPNKDKFCPTVNSLNLCKNOMSTWGS 1000
Db 858 AEDSDMETGKIMPNKNGSDGCVLAKVLVOTGSRDK--PIDISQLT--KORLVADAGI 913
Qy 1001 VAPSVQEPHKYL--PWFLND-----RPNKCPKGGLAAY--STSVNLTSDDQ 1043
Db 914 INPSA--FYIYLAWNSNDPVAAASQANIRHPRPFWMDK--ADYMETLIRIPAAEP 968
Qy 1044 VIASRFMAVHKPLKNSQDYTEALRAARELANITA-DLRKPGTDPAREVFYTTINVY 1102
Db 969 IEYAFPFPLNGLRDTSDEVEAIEKVRTICSNVTSGLSSYPNG-----YFP-----LPW 1018
Qy 1103 EGYLTLPBGLFMLSLCLVPFAVSCLLGLDLRSGLMLLSIVMLVDTVGFMALMDS 1162
Db 1019 EGYIGRHLPLFISVLAICTFLVCANFLNPTAGIT--VNVLAITVTLFGMGLIGIK 1077
Qy 1163 YNAVSLINLVASGVNSVEFVSHITRSF--AISTKPTMLERAKEATISMSAFAVAGVMTN 1220
Db 1078 LSAVPLVILIASGIVEFTVHALAFLAIGDKN--RAVLALBHMFAPIVLDG-AVST 1133
Qy 1221 LRGILVGLAKAQOLIOIFFRNLMLTLTGLHGLVFLVYIISYVC--PDVNPALALEO- 1277
Db 1134 LGLVLMLAGSEPDFYIRYFFAVLAITLIIIGVNLGLVLPVLLSFPGPYEVSFANGLANRL 1193
Qy 1278 --KRAEAAVAAVWVASCNHPKRSVSTADNIYVNSHSEGSIKG 1317
Db 1194 PTPSPPEPPSVVRFAMPQHTH--SGSDSSDSSEYSQTTVSG 1233

RESULT 11
US-09-724-631-19
Sequence 19, Application US/09724631
Patent No. 6551782
GENERAL INFORMATION:
APPLICANT: SCOTT, MATHEW P
GOODRICH, LISA V
JOHNSON, RONALD L
TITLE OF INVENTION: Patched Genes and their Use
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: US
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/724,631
FILING DATE: 28-NOV-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/656,055
FILING DATE: 1996-05-31
APPLICATION NUMBER: 08/540,406
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Rowland, Bertam I
REGISTRATION NUMBER: 20015
REFERENCE/DOCKET NUMBER: a60190-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 1447 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 19:
US-09-724-631-19

Query Match 9.4%; Score 651.5; DB 2; Length 1447;
Best Local Similarity 22.0%; Pred. NO. 1.1e-51;
Matches 300; Conservative 181; Mismatches 460; Indels 421; Gaps 47;

Qy 225 PQQAVSGIQLPNEVARNCSQGDVATCCQDCACAPAIARQALDSTFYLGOMGS 284
Db 24 PGRPAGGRRRRTGRLRAAARDRD-----YLHRSYCDAAFALEQI--- 65
Qy 285 LVLIILISVAVVITLLVGFRAVARADSKKAVDPK-----KGTSLSDKLSFSTHTLL 337
Db 66 -----SKGATGKAPLMYLRKQRLFLKIGCTIQKNC 98
Qy 338 GQFQGWGTWASWDLTVLVSIPVVALAAGLVETELTDPVELMSAPNSQARSEKAFH 397
Db 99 GKFE-----LVVGLLIFGAFVGLKANLETNVEBELWVEGVSRRLNYT 143
Qy 398 DQHFPPFRITN-QVLTLPNRSSTYRDSLLGPKNFSGILDLDLLELLE--LQERLRL 454
Db 144 RQKIEEAMFNPQMIQCPKKEG-----ANVLTTEALLQHLDSALQASRVHV 190
Qy 455 QVWSPDAQRNLSLDICYAP-----LNPDNSTLYDCCINSLLQYFQNNRTLLTLTA 505
Db 191 YKYN-----RQKLEHLCYKSGELITETGYMOQIIEYLPCLITPDLDFWEGAKLQSGTA 246
Qy 506 NOTLMGO-----TSQVDWKDHPLY-----CAN----- 527
Db 247 --YLLGKPLMTNTPDPLEBELKKIYQVDSWEMLNKAEGVGHGMDRCLNPADPDC 304
Qy 528 -----APL-----TFKDG-----ALALSCMA 544
Db 305 PATAFNKSTYKPLDALVINGGCHLSRKYMHMOBELIVGGIVKNSSTKLVSANHLQTMF 364
Qy 545 DYGAVPFPFLAIGVKGKDYSEABALIMTFGLNNVPADPPLAQKXKLEEAFLBEMRAFQ 604
Db 365 QLMTKQKMY--EHRKGYEY-----VSHIMWNE--DKAAALIEAMQRTYVEVHOSV 411
Qy 605 RMAQMFOVTFAESRLDEINRTAEDLPFATSYIVIFLYISIALGYSWSRWVDS 664
Db 412 AQNSIQKVLSPFTT--TTLDILKSPFSDVSVIVASGYLLMLVACLT--LRMD--CGKS 465
Qy 665 KATLGLGVAVVLGAVMAAMGFPSYLGIRSSLVILIQVFPVLVSYGADNIFLVLEYQRL 724
Db 466 QGAVGLAGVLLVLAISVAAAGLGLCSLIGISFNAATQVPFLALGVDVDFLAHAFST 525
Qy 725 PRRPPEPEVHIGRALGVAPSMMLCSLSEALCFELGLTTPMPAPRTALNSGLAVIIDF 784
Db 526 GGNKRIPEEDRTGECLKTGASVALTISINVTAFPMALIPBALRAESLOAAVVVNF 585

[illegible]

```

1  APPLICATION NUMBER: US/08/954,701A
2  FILING DATE: 20-OCT-1997
3  CLASSIFICATION: 435
4  ATTORNEY/AGENT INFORMATION:
5  NAME: Vincent, Matthew P
6  REGISTRATION NUMBER: 36709
7  REFERENCE/DOCKET NUMBER: SUV-003.080
8  TELECOMMUNICATION INFORMATION:
9  TELEPHONE: 617-832-1000
10 TELEFAX: 617-832-7000
11 INFORMATION FOR SEQ ID NO: 19:
12 SEQUENCE CHARACTERISTICS:
13 LENGTH: 1447 amino acids
14 TYPE: amino acid
15 STRANDEDNESS: single
16 TOPOLOGY: linear
17 MOLECULE TYPE: protein
18
19 US-08-954-701A-19

```

Query Match	9.4%	Score 651.5;	DB 2;	Length 1447;
Best Local Similarity	22.0%;	Pred. No. 1.1e-51;		
Matches 300;	Conservative 181;	Mismatches 460;	Indels 421;	Gaps 47;

225 PGQVGSGLPTNEGVARCNEGQDVATCSQDCAASCFAIRPQALDSTFLGOMGGS 28

285 LVLLIILCSVFAVVTILLVGFPRVAPARDKSKMNDPK-----KGTSLSDKLSFSTHTLL 3131

Db 66 -----SKGATGKAPLWLRKAFORLLFKLGCIQKNC 98

```
Db          ||::|||::|||:|:
99 GKF-----LVGLLFGAFVGLKANLETNVEELWTEVGSRVELNYT 143
```

```

398 DQHEGPFERTN-ÖVILTAPKRSSRYRDSLLGPKNSGIEDDLLLELE--LQERLRHL 454
      :::::  :::::

```

455 QWSPBQNRNLSLDDICYP-----LNPDTSLYDCCNSLLOYFÖNNRTLLLLTA 505

Db 191 YMYN---RQWKLEHLCKYSGELITETGYNDQIIEVLYPCLITTPDLCFWEAGAKLQSGTA 246

Db 247 --YLGKPPRIWTNFDPLEELBEELKKINYQVDSWEMLNKAEGHGCMYDRPCINPADEBC 304

```

QY      528 -----APL-----TFKDG-----ALALSCMA 544
          ||              ||: ||

```

545 DYGA PVPPELAI GSYKGDSEAEALIMTESLNNYPAGDPLAQA KIMEEAFIEEMRAFO 6044

Db 365 QLMTPKOMY---EHFKGYEY-----VSHINWNE-----DKAAALLENMORTYVEVHQSV 4111

[illegible]

665 KATIGLGVAVVLGAVMAANGFFSLIGIRSSLVILQVPPVLTVSGADNIFITVLEYORL 724

725 BBDDCEDEEVHTGAI,CPVADSM,I,CSI,SEA,TCEFI,GA,TQMDAVTEAI,TSGI,AVT,INF 7848
DB 466 QGAVGACVLLVAUSVAAAGGGLCSLIGISFNMAI,IOV,PF,ILALGVGVDF,FLMATA,GBI 323

Db 526 GÖNKRİPFEDRTGECLKRTGASVALTİSİNTAFTFMAALİFİPALNAFSLÖAAVVVVNF 565

[illegible]

823 QGBGL----- 827

Db 646 SSHAHEQTITMSTVOLRTEYDHTHYTTTAEBSRSEISVQPTVTOTDITLSCSPBST 705
Qy 828 -----LIGFQKAVAPFLHWTIRGVLLFLALFGVSL 861
Db 706 SSTRDLDSQFSDSLHCLERPCCKWTLSFAEKHAPFLKPKAKVVIFFLGLIGVSL 765
Qy 862 YSMCHI SVGLDDELALPKDSYLLDYFLNRYFEVGAVPYVTTLTGYNSSSAGNAIACS 921
Db 766 YGTTVRDGLDLDIVPRETREVDFIAQFKYFSF-----YNN----- 803
Qy 922 SAGCNNSFTOKIOYAT-----EFPEOSYLAIPASS-----WVDDFIDWL----- 961
Db 804 -----YIVTQKADYPNIOHLLYDLHRSFNKTYMLEENKQLPKMWLHYFRDMIGLOD 857
Qy 962 -----TPSS-----CCRLYISGPNKDKFCSTVNSLNCNKMSITWGS 1000
Db 856 AFDSMETGKIMPNNYKNGSDGVLAAYKLVTGSRDK--PIDISQLT--KQRLVDADGI 913
Qy 1001 VAPSVEQPHKYL--PWFLND-----RPNIKCPKGGLAAY--STSVNLTSDGQ 1043
Db 914 INPSA--FYIYLTAVWSNDPVAYVAAQANIRPHRPEWVHK--ADYMBETRLRIPAABP 968
Qy 1044 VLASRFMAVHKPLKNSODYTEALRAARELANITA--DLRKVPGTDPAPFVPPYTTINVFY 1102
Db 968 IEYAQPFYLNGLRDTSDVEAIEKVRTICSNYTSIGLSSYPNG-----YFP-----LFW 1018
Qy 1103 EOYLTILPEGLFMLSCLVPTFAVSCLLGLDLRSGLLNLLSIWMILVDVTFGMALMDIS 1162
Db 1019 EGYIGLRHMLLFISIVLACTFLVCVFLNMPWTAGII--VMVLALMTVELFGMGLIGIX 1077
Qy 1163 YNAVSLINLVSAVGSVEFVSHITRSF--AISTKPTWLERAKAETISGSAVPAVAMTN 1220
Db 1078 ISAAPVLIILASVIGVEFTVHVALAFLTAIDKN--BRAVALAHEHMAFAPLVDG--AVST 1133
Qy 1221 LEGIIVLGLAKQOLQIIFPFRNLITTLGLHGLFVLISYVYG--PDVNPALALBQ- 1277
Db 1134 LAGVLMLAGSEDFIVRYFAVALAITTIGVNLGLVLLVPLSFPGRPYEVSFPAINGLRL 1193
Qy 1278 --KRAEAVAAVMVASCPNHPRSVSTADNIYVNHSPFSGIKG 1317
Db 1194 PTPSPPEPPSVVRPAMPQHTH--SGSDSSDSEYSQTTVSG 1233

RESULT 13
US-09-807-007-6
; Sequence 6, Application US/09807007
; Patent No. 6881833
; GENERAL INFORMATION:
; APPLICANT: ZAPHIROPOULOS, Peter et al.
; TITLE OF INVENTION: A NOVEL COMPONENT IN THE HEDGEHOG SIGNALLING PATHWAY
; FILE REFERENCE: 2921-0130P
; CURRENT APPLICATION NUMBER: US/09/807,007
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 6
; LENGTH: 1447
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-807-007-6

Query Match 9.4%; Score 651.5; DB 2; Length 1447;
Best Local Similarity 22.0%; Pred. No. 1.1e-51;
Matches 300; Conservative 181; Mismatches 460; Indels 421; Gaps 47;

Qy 225 PQAGVSGGIQPLNEGARGNESGDDVATACSCODCAAPARAPALDSTFYLGMPSS 284
Db 24 PERPAGGGRRRTGTGIRRAAPDRD-----YLRPSYCDAAFALEQI--- 65
Qy 285 LVLIILICSVFAVVTILLVGFVAPARADSKMVDPK-----KGTSLSDKLSFTHTLL 337
Db 66 -----SKGKATGRAPLMLRAKFPRLFKIGCYIQKNC 98

Qy 338 GQFPGMGKGTWASWPLTLLVSVIPVALLAGLVFTELTTPPVLEMSAPNSQABSEKAFH 397
Db 99 GKF-----LVVGLLFGAFVAGLKAANLETWBLEWGVRSRELNYT 143
Qy 398 DQHFGRFRTN--QVILTAPNRSRYRDSLLGPKNFSGILLDLLELLE--LOERLRL 454
Db 144 ROKIGSEAMFPQMLIQPKKEG-----ANVLTEALLQHLDSALQASRHHV 190
Qy 455 QVMSPEAQRNITSLODICYAP-----LNPNTSLYDCCINSLLQYFQNNRTLLLLTA 505
Db 191 YMYN-----RQWLEHLCYKSGELITETGYMDQIIEIYLPCLITPLDCFMGAKLQSGTA 246
Qy 506 NOTLNGQ-----TSQYDKMDHFLY-----CAN----- 527
Db 247 --YLLGKPEPLRWTFPDLFLEELKKNYQVDSWEMLNKAEGHGYMDRCLNPADDC 304
Qy 528 -----APL-----TPKQGT-----ALALSCMA 544
Db 305 PATAPNKNSTKPLDMALVNLGCGHGLSKRYHMQBELIVGTVKNSTGKLSAHLQTMF 364
Qy 545 DYGAIVFPFLAIGGYKDYSEAEALIMTFSLNNYPADPRLAQAQKWEAFLEEMRAFO 604
Db 365 QLMTEKQWY---EHRKGYEY-----VSHINME---DKAAALILEAQRTYVEVHQSV 411
Qy 605 RRMAGMFOVTPFAERSLEDEINRTTAEDLPFAFSYIYIFLYISLAGSYSGWSRWVDS 664
Db 412 AONSTQKVLFTT--TTLDLILKSPSDSVIVASGIVLMLLVACTM--LRMD--CSKS 465
Qy 665 KATLGLGVAVVVAGVMAAMFPFSLYGRSSLVLIQVVPVLVSGADNIEFLVLEYORL 724
Db 466 QCAVLAGVLVVALSVAAGLGLCSLIGISFMAATTOVLPFLALGCVUDVFLMAHAFET 525
Qy 725 PRRPEPREVHIGRALGVAPASMLLSLSEALICEFLGATMPAPVTFALTSGLAVIDF 784
Db 526 GQNKRIIPREDRTGECTKATGASVALTISNTAFPMALIGIPALRAFSLQAAVVVWVF 585
Qy 785 LLQMSAPFALLSLSKQREASRLDYCC-----VNRQEL-----PRG 822
Db 586 AMVLLIFPAILISMDLYRREDRLDFCCFTSPCVSRVIVQBPQAYTDTHTNTRYSPPPY 645
Qy 823 QGEGE----- 827
Db 646 SSHAHEQTITMSTVOLRTEYDHTHYTTTAEBSRSEISVQPTVTOTDITLSCSPBST 705
Qy 828 -----LIGFQKAVAPFLHWTIRGVLLFLALFGVSL 861
Db 706 SSTRDLDSQFSDSLHCLERPCCKWTLSFAEKHAPFLKPKAKVVIFFLGLIGVSL 765
Qy 862 YSMCHI SVGLDDELALPKDSYLLDYFLNRYFEVGAVPYVTTLTGYNSSSAGNAIACS 921
Db 766 YGTTVRDGLDLDIVPRETREVDFIAQFKYFSF-----YNN----- 803
Qy 922 SAGCNNSFTOKIOYAT-----EFPEOSYLAIPASS-----WVDDFIDWL----- 961
Db 804 -----YIVTQKADYPNIOHLLYDLHRSFNKTYMLEENKQLPKMWLHYFRDMIGLOD 857
Qy 962 -----TPSS-----CCRLYISGPNKDKFCSTVNSLNCNKMSITWGS 1000
Db 856 AFDSMETGKIMPNNYKNGSDGVLAAYKLVTGSRDK--PIDISQLT--KQRLVDADGI 913
Qy 1001 VAPSVEQPHKYL--PWFLND-----RPNIKCPKGGLAAY--STSVNLTSDGQ 1043
Db 914 INPSA--FYIYLTAVWSNDPVAYVAAQANIRPHRPEWVHK--ADYMBETRLRIPAABP 968
Qy 1044 VLASRFMAVHKPLKNSODYTEALRAARELANITA--DLRKVPGTDPAPFVPPYTTINVFY 1102
Db 968 IEYAQPFYLNGLRDTSDVEAIEKVRTICSNYTSIGLSSYPNG-----YFP-----LFW 1018
Qy 1103 EOYLTILPEGLFMLSCLVPTFAVSCLLGLDLRSGLLNLLSIWMILVDVTFGMALMDIS 1162
Db 1019 EGYIGLRHMLLFISIVLACTFLVCVFLNMPWTAGII--VMVLALMTVELFGMGLIGIX 1077
Qy 1163 YNAVSLINLVSAVGSVEFVSHITRSF--AISTKPTWLERAKAETISGSAVPAVAMTN 1220

Db 1078 ISAAPVILIASGIGVEFTVHVALAFLTAIDGN---RRAVLALEHMFAPVLGD-AVST 1133
Qy 1221 LGGILVLGAKQLOIOPFFRLNLITLGLLHGLVFLPVILSYVG--PDVNPALALBQ- 1277
Db 1134 LGLVLMASBDFPIRYFAVALITLIGLVALLPLVLSFFEPYFEVSPANGLNRL 1193
Qy 1278 --KRAEAAVAVWVASCNNPSPSVSTADNIYVNHSPESIGK 1317
Db 1194 PTPSPPEPPSVVRFPMPFGHTH--SGSDSDSEYSSQTIVSG 1233

RESULT 14
US-09-754-032-19
Sequence 19, Application US/09754032
Patent No. 6921646
GENERAL INFORMATION:
APPLICANT: SCOTT, MATHEW P
GOODRICH, LISA V
JOHNSON, RONALD L
TITLE OF INVENTION: Patched Genes and their Use
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: US
ZIP: 94111

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/754,032
FILING DATE: 03-Jan-2001
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/540,406
FILING DATE: 06-Oct-1995
ATTORNEY/AGENT INFORMATION:
NAME: Rowland, Bertram I
REGISTRATION NUMBER: 20015
REFERENCE/DOCKET NUMBER: A60190-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 1447 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 19:
US-09-754-032-19

Query Match 9.4%; Score 651.5; DB 2; Length 1447;
Best Local Similarity 22.0%; Pred. No. 1.1e-51;
Matches 300; Conservative 181; Mismatches 460; Indels 421; Gaps 47;

Qy 225 PGAVSGGIQPLNMGVARGNESQGDVATCSQDCASCPAIAAPQALDSTFYLGWPGS 284
Db 24 PGRPAGGGRRRRGGRLRAAAPDRD-----YLRPSYCDAAFALEQ--- 65
Qy 285 LVLIILICSVFAVVTILLVGRVAPARDKSKMVPK-----KGTSLSDKLSFSTHTLL 337
Db 66 -----SKGKATGRKAPLMLRAKFGQLRLKGLGCIQKNC 98
Qy 338 GQFPQMGKTVWVWMLTILVSVIVVLAAGLVFTLTDPVLEMSAPNSQARSECAFH 397
Db 99 GKF-----LVVGLLIFGAFVGLKAAMLTINVELMWEVGVGRVSRRLNNT 143

Qy 398 DQHFQFFRITN-QVITLTPANSSRYRDSLLGKPNFSGILDLDLLELE--LQERLRLH 454
Db 144 RQKIEEAMFNQOLMIQPKERG-----ANVLTEALLQHLDSALQASRVH 190
Qy 455 QVWSPBQRNISLQDICAAP-----LINDNTSLVDCINSLSLOFQONRTLLTLTA 505
Db 191 YNIN-----RQWLEHLQYKSGELITETGYMDIILEYLPCLLITPLDCTWBAKIQSGTA 246
Qy 506 NQTLNGQ-----TSQVMDKHFLY-----CAN----- 527
Db 247 --YLLGKPPRLRWNTDPLFLEELKKINQVDSMEEMKAEVGGYMDRPLCLNPRDPC 304
Qy 528 -----APL-----TFKQGT-----ALALSCNA 544
Db 305 PATAENKSTKPLDMLVINGCHGLSKRYMMQBELIVGVVKNSTGLVSAHMLQTMF 364
Qy 545 DYGAVPFPLAIGYKGDYSEAEALIMTFSLINNYPAQDPLAQAQKMEAFLEKRAFO 604
Db 365 QLMTPKQWY---EHFQYGY-----VSHINNE---DKAAAILQAMQRTYVEVHQGV 411
Qy 605 RRMAGMFOVTFPAERSLEDEINRTTAEDLPATSYIVFLYISLALSGYSSMSRMVDS 664
Db 412 AQNSTQKVLSTFTT--TLLDDILKSFSDVSIVRVASGILLMLAYACLTM--LRMD--CSKS 465
Qy 665 KATLGLGVAVVLGAVMAAMGFSYLGTRSSLVILQVPELVISGADNIFIVLEYQRL 724
Db 466 QGAVGLAGVLLVALSVAGLGLCSLIGSENAATTQVPLALGVGDVDFLAAHAFSET 525
Qy 725 PRPGPREVHIGRALGRVAPSMCLSLSEALCFIFGALTTPMPAVRTFALTSGLAVLDF 784
Db 526 GQNKRIPEFDRIGECIKRTGASVALTSISNVAFFMAALIPALRAFSLQAAVVVVFNF 585
Qy 785 LLOMSAFVALLSLDSKROEASRLDVCCC-----VKQDEL-----PPG 822
Db 586 ANVLIFPALISMDLYRREDRLDIFCCTSPCVSRVIOVEQATYDTHDNTRYSPPPY 645
Qy 823 QGEGEL----- 827
Db 646 SSHSFAHETQITMGSTVOLRTEYDPTHVYTTAEPRSEISVQPVTVTQDTLSCQSEST 705
Qy 828 -----LGFPOKAVAPFLMLHITRGVLLPLALFGVLL 861
Db 706 SSTRDLLSQFSDSLHCLBPCTKWLTSFPAKHVAPFLKPKAKVAVIFLFGILGVSIL 765
Qy 862 YSKCHISVGLDQELAPKDSYLLDYFLINRFEVAVPYFTYTLGYNSESSEGMALIS 921
Db 766 YGTTVRVDGLDLTDIVPRETRBYDFIAQKTFYSF-----YNN----- 803
Qy 922 SAGCNNPFTQKIQVAT-----EPPEGSYLAIPASS-----WVDFIDML----- 961
Db 804 -----YIVTQKADYVNIQHLVYLHRSFSNVKTYVMLEENKQLPKMMLHYFRDMLQGLD 857
Qy 962 -----FPSS-----CCRLYISGPNKDFCPTVNSLCLKNKMSITMGS 1000
Db 858 APDSMETGKIMPNNYKNSDDGVLAAYKLIVOTGSDK--PIDISGLT--KQRLVADAGI 913
Qy 1001 VRPSVEQFHXYL-PWFLND-----RPNIKCPKGLAAY--STSVNLTSQGO 1043
Db 914 INPSA--FYIYLTAWSNDPVVAYASQANIRPRHPWVHK--ADYMPETRLRIIPAAB 968
Qy 1044 VLASRFMAHYKPKNSQDYTEALRAARELANITA-DLAKVPETDADFVFPFTITNVF 1102
Db 969 IEVAGFPFYINGLRDTSDFEALAEKRTTICSNYTSGLSSYPNG-----YFP-----LFW 1018
Qy 1103 EQVLTILPBGFLPMLSLCYLPTFAVSCLLGLDRLSGMLMLSTGLVLPVGVFMALMDS 1162
Db 1019 EGYIGLRHMLLLEFISVLACTFLVCAVFLNPTAGIT--VMVALMTVELFGMGLIGLK 1077
Qy 1163 YNAVSLINIVSAVGMVEFVSHITRSF--AISTKPTMLBRRAKATISMGSAVPAGVAMTN 1220
Db 1078 LSAVPAVILIASGIGVEFTVHVALAFLTAIDGN---RRAVLALEHMFAPVLGD-AVST 1133

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using bw model

Run on: March 22, 2006, 22:51:02 ; Search time 53 Seconds
(without alignments)

2418.127 Million cell updates/sec

Title: US-10-736-769-4

Perfect score: 6909

Sequence: 1 MAEAGLRGWLWALLRLRLAQ.....GSIKGAGAIENPLPNNGRQF 1332

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2385	34.5	1278	2 T30188	Niemann-Pick C dis
2	1394	20.2	1055	2 T05663	hypothetical prote
3	1347	19.5	1456	2 T15961	hypothetical prote
4	1329	19.2	1170	2 S52525	probable membrane
5	646.5	9.4	1434	2 T30172	transmembrane prot
6	628.5	9.1	1442	2 T18538	patched protein -
7	600	8.7	915	2 S44797	F0968.4 protein -
8	598.5	8.7	933	2 T25600	hypothetical prote
9	583.5	8.4	1182	2 T13952	membrane protein p
10	573.5	8.3	1003	2 T26746	hypothetical prote
11	572	8.3	1220	2 T18291	patched protein -
12	539.5	7.8	889	2 T29590	hypothetical prote
13	514	7.4	1405	2 T27969	hypothetical prote
14	473.5	6.9	1015	2 T15930	hypothetical prote
15	464.5	6.7	1299	1 S06119	membrane protein p
16	458.5	6.6	955	2 T21612	hypothetical prote
17	436	6.3	952	2 A89153	protein C24B5.3 [1
18	409.5	5.9	936	2 T26521	hypothetical prote
19	407	5.9	800	2 T26683	hypothetical prote
20	399.5	5.8	881	2 T31739	hypothetical prote
21	381	5.5	840	2 T33217	hypothetical prote
22	379	5.5	714	2 T16126	hypothetical prote
23	367	5.3	690	2 T23399	hypothetical prote
24	367	5.3	877	2 T24097	hypothetical prote
25	349	5.0	845	2 T25657	hypothetical prote
26	347.5	5.0	820	2 T32908	hypothetical prote
27	312	4.5	890	2 T22186	hypothetical prote
28	297.5	4.3	871	2 T28706	hypothetical prote
29	294	4.3	983	2 T21213	hypothetical prote

30	276.5	4.0	633	2 S44795	F0968.3 protein -
31	259.5	3.8	413	2 S28276	hypothetical prote
32	224	3.2	1276	2 T18526	SREBP cleavage act
33	189.5	2.7	1154	2 T48829	related to SREBP c
34	187.5	2.7	932	2 T28820	hypothetical prote
35	186.5	2.7	1227	2 T20370	hypothetical prote
36	179.5	2.6	787	2 H71453	hypothetical prote
37	173	2.5	969	2 T33156	hypothetical prote
38	170	2.5	746	2 A75018	transport protein
39	154	2.2	823	2 B81282	probable integral
40	149	2.2	1011	2 T07712	probable ABC-type
41	149	2.2	1051	2 A93455	multidrug resistanc
42	148.5	2.1	746	2 H84301	hypothetical prote
43	146	2.1	1041	2 AC0423	multidrug efflux p
44	143	2.1	724	2 H69780	antibiotic transpo
45	141.5	2.0	1049	2 AF0561	acriflavin resist

ALIGNMENTS

RESULT 1

T30188
Niemann-Pick C disease protein - mouse
N:Alternate names: NCP1 protein
C:Species: Mus musculus (house mouse)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T30188
R:Loftus, S.K.; Morris, J.A.; Carstea, E.D.; Gu, J.Z.; Cummings, C.; Brown, A.; Ellison, Science 277, 232-235, 1997
A:Title: Murine model of niemann-pick C disease: mutation in a cholesterol homeostasis A:Reference number: Z20765; MUID:97362324; PMID:9211850
A:Accession: T30188
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1278 <LOF>
A:Cross-references: UNIPROT:Q35604; UNIPARC:UP1000002995C; EMBL:AF003348; NID:G2251241; C:Genetics:
A:Gene: NPC1
A:Map position: 18

Query Match 34.5%; Score 2385; DB 2; Length 1278;
Best Local Similarity 40.2%; Pred. No. 1.2e-158;
Matches 525; Conservative 233; Mismatches 443; Indels 104; Gaps 27;

QY	14	LLRLAQSPPYTHIQPGICAFYDCCGKNPBLSSGLMTLSNVSCLSNTPARKITGDHLTL	73
DB	12	LLRLCPAQVFSQ-----SCVMYGGCG--IATGD--KRYNCKYSGPKPLPKDGYDL	59
QY	74	LQKICPRLVTGPNTOACSAKQVLSLEASLSTTKALTTRCPACSDNFVNIHGNTCSPNQ	133
DB	60	VEELCPGLFF-DNVLSCDIIQDLQTKSNLQPLQPLSLSCPCSFYNIIMTLPELITCSPHQ	118
QY	134	SLFINVTRVAQAGQLPA---VVAEAFYQHSFPAEYSDSCSRVRVPAALATLAVTNC	189
DB	119	SGFLNVTATEDYEDFPKTPENKTNKELEYVVGCSFANAMNYACRDVEASSNEKALGLIC	178
QY	190	GYVGSALCNAQRMWNFQSGTNGGLAPLDI-----TFHLLPEQNAVSGIQPLEGVARCN	244
DB	179	GDARA-CNATWIEFMKDKNGQAPFTIIPVFSDLSTI-----GMEPMNATKCN	229
QY	245	ESQGDVATSCCODCAASC-----PAIRPOLADSTF-----YLGOMGSLV	286
DB	230	ESYDEVTPGSCODGSIYCGPRPQPPPPMPRIRINGLDMMYIMVTVYAFIFVFFGALL	289
QY	287	LIIILCSFAVVTILLVGFVAPARDKSRVDPKGTSLSDYLSFSTHTLLQFPQSGCT	346
DB	290	AWCHRRRYFVSEYTPIDENIAFVNS--DKGEASCCDPLGAFFDCLRMFTKMA	345
QY	347	WASNPFLTLVLSVIRPVALLAGLVFTTELTTPVLELWSAPNSQARSEKAFHOHGGPR	406
DB	346	FCVRNPCTIIFPSLAFITVCGSLVAVQVTVNPVELMSAPHQARLEKEFKHGFPR	405

QY 407 TNOVLTAPNRSSRYRDSLLELPK-NFSGILLDILLELELOERLRHQVSPENQR- 464
DB 406 TEOLIIQAENTSHVIEPYRAGADVFCFPLNKEIILHQVLIHQ-----IAIESITASYNN 460
QY 465 --ISLQDICAPLNPNTSLVYOCCTISLQYFONNRTLLLTANQTLMGQTSQV----- 516
DB 461 EYVTLQDLCVAPLSPYNN--NCTIWSVNTYFONSHAVL-----DSQVGDFFYI 506
QY 517 --DWKDHFLYCANAPLTFKDGTAALASCMADYGAPVPEPLAIGYKGDYSEALIMTF 574
DB 507 YADYHNFHLYCVRAPASLNDTSLHGPCGTGCGPFPMLVYGVDQDQYNNATALVITF 566
QY 575 SLANTPAGPRLAQAOLMEAEFLIEBKRAQORMAGFOYTFTRERLEDEIRTTAEDLP 634
DB 567 PANNYNTDERLQRAWAMEKEFISFYKNYKN--PMLTISFPAERSIEDELRESNDYF 623
QY 635 IFATSYIVIFLYISLALGSYSSWRVWDSKATIGAGVAVVLGAVMAMGFSSYIGIRS 694
DB 624 TYIISYVMFLYISLALGHIOGSRLLVDSKISLGLAGILIYLSVACSLGIFSYMGML 683
QY 695 SLVILQVNPFLVLSVGADNIFIVLEEQRLPRRPGSPREVIHGRALGRVAPSMILCSLSE 754
DB 684 TLIVIEVIFPLVAVGVNDIFILVQYQDERLQEBTLDOQLGRILGEVAPTFMILSSFSE 743
QY 755 AICFPLGALTMPAYARTFALTSGLAVITLFLQMSAFVALLSDSKROBASRLDVCCKK 814
DB 744 TSAFFFGALSMYAVHTFSLFAGMAVLIDFLQITCFVSLGLDTRKQKRNHLDILCCVR 803
QY 815 POELPPPGOG----EGLLGPFQOKAYAPFLMHITRGVLLFLFALFGVSLXSMCHISVG 870
DB 804 GAD---DGGSHASESYLFRFFKNYFAPLLKMDLRPIYAVFVGLSVAVAKVDIG 860
QY 871 LDQELALPKDSYLLDYFLFLNRYFEVGAPVYFVTLTGYNFSSBAGNNAICSSAGCINPFSF 930
DB 861 LDQSLMPDYSYVIANFSLAQLHSGRPVYFVLEEGYNSYSSKQGMVCGMGCONDBL 920
QY 931 TOKIQYATFEPESQSYALTPASSWVDFIMWLT-SSCCRLYISGPKKXKFCSTYNSLNC 989
DB 921 VOQIFMAELDYTRTGFAPSSWIDYFPMWSFQSSCCRLY--NVTNHFCCASVWDPTC 977
QY 990 LKNCMSIT-MGSVRPVEQFHKYLPWFLNDRPNIKPKGGLAAYSTSVNLTSQGYLT-AS 1047
DB 978 VR-CRPLTPEGKORPGKGFPMKFLPMFLSDNPNPKCKGKGAHAYSAVIVGDDTYTGAT 1036
QY 1048 RMAVYHKPLKNSQDYTEALRAARELANITADLRKVPGTDPAFEVFPYITTVFEQYLT 1107
DB 1037 YEMTYHTILKTSADYTDAMKAKRLIASNITETWRS-KGSD--YRVPFYSVFVFEQYLT 1093
QY 1108 ILPEGLFMLSCLVFPFANVSCLLGLDGLASGLNLSTIMVILVDYVGPALNDISTYNS 1167
DB 1094 IIDDITIFNLVSLSIGSIFLVTLVVLGCELSAVIMCITIMILVNMFGVMLGSLNAVS 1153
QY 1166 LNLVSAVMSGYEVEFVHITRSPALSTKPTMLERAKATISMGSAVPAVAMTNLPGLIYL 1227
DB 1154 LNLVMSGISIVBFCHITRAFTMSYKSGRVSRAEBALAHMSSSVSGITTLTKFGGIIVL 1213
QY 1228 GLAKAQLIQIFPFRILNLITLLGLLGLVFLPYILSYVDPVNP 1272
DB 1214 ARAKQDIFEIFYFRMYLNAVLLGATHTGLIFLPVLSYIGPSVKA 1258

RESULT 2
T05663
hypothetical protein F22113.120 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 09-Jul-2004
C:Accession: T05663
R:Bevan, M.; Wedler, H.; Kutzner, M.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Mayer, K.R.
submitted to the Protein Sequence Database, February 1999
A:Reference number: Z15420
A:Accession: T05663
A:Molecule type: DNA
A:Residues: 1-1055 <BBV>

A:Cross-references: UNIPROT:O95VF0; UNIPARC:UPI00000A552C; EMBL:AL035539
A:Experimental source: cultivar Columbia; BAC clone F22113
C:Genetics:
A:Map position: 4
A:Introns: 24/3; 60/3; 99/3; 150/1; 193/1; 216/3; 278/3; 297/2; 336/3; 364/3; 396/3; 41
026/3
A>Note: F22113.120

Query Match 20.2%; Score 1394; DB 2; Length 1055;
Best Local Similarity 28.5%; Pred. No. 2.6e-89;
Matches 377; Conservative 221; Mismatches 385; Indels 340; Gaps 41;

QY 36 YDECGKNEPLSGSLMTLSNGLSTNPARKITGDHLI--LLOKICPRLYGTENQACCSA 93
DB 2 YDICHHRSD-GKVL-----NCPYASPS--IQDELFSAKIQSLCPTI---SGNVCTE 48
QY 94 KQLVLEASLSTIKKLLTRCPACSDNPNLHCHNCSNOSLFIVNTRVAOLGAGQLPAY 153
DB 49 TQFDLRSQVQOAVPFLVGCRCACLRNFILNLCESCSFNOSLFIVNTRVAEV-SCNL-TV 106
QY 154 VAYEAFYQHSFAQSYDSCSRVRVPAATLAVGTWGYGALCNQRMNFGD--TG 210
DB 107 DQIDYHITDTRGEGLYESCKEYKFGTMNTRAINFV---GGAKNFRMFETFIGQKASG 162
QY 211 NGLAPLDITFHLEFGQAVSGCIQPLNCGVARNCHESQDDVATGSCQCCAAACRAIARQ 270
DB 163 FPGSPYAINFSSSLP---ESSAMVFN-----VSYSACSSPEPLP 203
QY 271 ALDS-TFYLG-----OMPSLVLIILIGSVFAVNTLLVGFRAVAPARDKSKMDPKK 322
DB 204 DEDSCSITIGPLKRCIELSMALYVLLVSCFPGAGNRRNRTQPLDSSKPLHPBE 263
QY 323 TSLSDKLSFSTHTLLG-----QFQCGMTVNASWPLTILVSVIPVA 365
DB 264 DQINSEMK---ENILGVKVRHQAQLSPQRYMAKYRSYSGWIMANPFLVFMVAIVLA 320
QY 366 LAAGVFLPBLTDPPELMSAPNSQARSKAHQDFGFPFRTNYIL-TANRRSYRDS 424
DB 321 LCGSYNKRKVEFRPKLWVGESKAEKKEFDTHLSPFYRIEOLILATVDDPKSGRPS 380
QY 425 LLLGPNFSGILLDLLELELOERLRHLQWSPDQARNISLQDICAPLNPNTSLYD 484
DB 381 I-----VTDEIIL-LPDIQK----- 396
QY 485 CCINSLQYFONNRTLLLTANQTLMGQTSQVDMKHFLYCANADLTFKDGTALASGMA 544
DB 397 -----YFKMD-----SGTFDDYGVGEHAEC-----FOHYTS-SETCLS 429
QY 545 DYGAVPFPLAIGYKGDYSE-----ABALIMTFSLNYP 580
DB 430 AFQAEVVDPSAVLGGFSGNNYSBWVSELGCVSPDCSDVKTLEFOATAFVVTYVNNV- 488
QY 581 AGDP--RLAQAOLMEAEFL---EEMRAFORRMAGFOVTFTRERLEDEIRTTAEDLP 634
DB 489 IGDSSNENAKRAVAMKSRITQLAKEBLPMVR--SKNLSLSTSSBSIESBLKREBTAIV 546
QY 635 IFATSYIVIFLYISLALGSYSSWRVWDSKATIGAGVAVVLGAVMAMGFSSYIGIRS 694
DB 547 TIAASYLVWFYIVSTLDADAPQYTFYISSKVLGLSGVVLVLSVLSGVFASLGYKS 606
QY 695 SLVILQVNPFLVLSVGADNIFIVLEEQRLPRRPGSPREVIHGRALGRVAPSMILCSLSE 754
DB 607 TLIMEVIFPLVAVGVNDIFILVQYQDERLQEBTLDOQLGRILGEVAPTFMILSSFSE 743
QY 755 AICFPLGALTMPAYARTFALTSGLAVITLFLQMSAFVALLSDSKROBASRLDVCCKK 814
DB 666 VLAEPVAGFVPMFACRITSMFPAALAIMDFLQITAFVALLVFDCKSADRIRIDCFPLCK 724
QY 815 ----POELPPPGOGSGLLIGFQOKAYAPFLMHITRGVLLFLFALFGVSLXSMCHISVG 870
DB 725 VPSSRSVEGGRREGFLERVYKVEVHAVVLGLMGVKNVAVVAFV----- 769
QY 871 LDQELALPKDSYLLDYFLFLNRYFEVGAPVYFVTLTGYNFSSBAGNNAICSSAGCINPFSF 930

Db 770 -----FAL 772
QY 931 TOKIYATEPPEOSYLAIIPASSWVDDPFIIDLTPSS--CCRLYISGPNKOKFCSPSTVNSLN 988
Db 773 ASLISPASQASDSTSYIAKAPASMDPLVWLSPEAGCCCKFTNG-----SYCP---DDQ 825
QY 989 CLKNCMSITWGSVPSVEQFHXYLPWFLNDRPNKICPKGGLAAYSTSVNLT--SDGVOLA 1046
Db 826 CFEHS---DLVQRPSTAGREKLPWFNLNLPADCAKGGHGAVTNSVDLKGVESGVIQA 882
QY 1047 SRPMAYHKPKXSQDYTEALRARELANITADLRKVPGRDPAEVPPPTITNVFEQYL 1106
Db 883 SEERTHTPL-NTQ-----IDIFPSVFYIFFEQYL 912
QY 1107 TLPEGLFMLSCLVLPFAVSCLLGLDLSGLLNLSIWMILVDTVGFMAWDISYNNV 1166
Db 913 NIWTVALLTNLAIV-----GIQLANV 934
QY 1167 SLINLVSAGMSVEFVSHITRSPASTKPTWLERAKATISMGSAVAGVAMTNLPQILV 1226
Db 935 SYVNLMSIGIAVEFCVHISHATLMSGDR-BHRARBLETWGAASFSGITTLKLVGIV 993
QY 1227 LGLAKQQLQIFFFRNLTLTLGLLHGLVFLPYIISYVG-PDVNPALAEQRAEANA 1285
Db 994 LCFARSEIFVYVYFCMYLALVILIGFLHGLVFLPYIISLAGPPQLN--LDIEQOOTDEAS 1051
QY 1286 AVN 1288
Db 1052 SLL 1054

RESULT 3
T15961
hypothetical protein F02B8.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_rev1510n 20-Sep-1999 #text_change 09-Jul-2004
C:Accession: T15961
R:Miller, N.
submitted to the EMBL Data Library, April 1996
A:Description: The sequence of C. elegans cosmid F02B8.
A:Reference number: Z18436
A:Accession: T15961
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1456 <M11>
A:Cross-references: UNIPROT:Q19127; UNIPARC:UPI000017B8AD; EMBL:U53340; NID:91255859; PI
A:Experimental source: strain Bristol N2; clone F02B8
C:Genetic9:
A:Gene: CRSP:F02B8.6
A:Map position: X
A:introns: 71/1; 124/3; 218/3; 269/3; 310/3; 403/2; 484/3; 528/2; 764/1; 858/3; 888/3; 1

Query Match 19.5%; Score 1347; DB 2; Length 1456;
Best Local Similarity 27.5%; Pred. No. 7.7e-86;
Matches 372; Conservative 258; Mismatches 565; Indels 158; Gaps 41;
QY 76 KICPRLYTGPNYACCSAKQVLSLESLSTTKALLTRCPACSNFPNHLGHNTCSNOSL 135
Db 63 EFCPHLLTGDN-KLCTPSQAEGLTKQIQARHILRCPCFCDFNFAKLWCEFTCSPOOD 121
QY 136 FINTVVAOL--GAGQLPAVVAEAF-----YQHS--FAEQSYDSCSRVVPAAATLAVG 186
Db 122 FVSIEMKPIKEKGETPEYQPAEAYVNYEYRLSTDFABGMFSSCDYTFGGQPALRV- 180
QY 187 TWCYVGSALCNARWMLNFOGDTGNGL-APLDITFHLLBPGQAVSGIOP-LNEGVARCN 244
Db 181 -WC---TSFPCCTLNMLLEFIGTQNLDLNIPRHTKFLLYDPIKTPPSDRSTYVMNVFTGCD 236
QY 245 ESGGDDVATCSCODCAASCAPARFQALDSTFTLQGM-----PGLVLLIIILCSVAVVT 299
Db 237 KSARVGMPCSTSEC--NKEEYANLIDLDDGKTSGGTGNVHGIACLNIFVMAFIQSLAV 294

QY 300 ILLVGF-----RVAPARDXSKWVDPKKGTSLSDKLSFSTHTLLGQSF 341
Db 295 LILCVFVFTSYDEYDTNLRQTOGSESPKRNKIK---RTGAMI-----HNFEMENNA 342
QY 342 QGMGTWVAWMLTLLIVLSIIPVVALAAGLVFELTTPDPELMSAPNSQARSEKAFHDHF 401
Db 343 RDIQMGAGNPNSHFTIGCAVILPCLPGMTYHSESTNVVDMSSPPSRAROEEMVNNANF 402
QY 402 GPFPRITQVILTAPNKSSRYSDSLLPKPNFSGILDLDBLLELLEQERLRLHQVNSPEA 461
Db 403 GRPQRVQIML--SHRDFQSSGKLXGF-----VFHKDIFELFDILNMIKNISTQDSOG 455
QY 462 QRNISLQDICYAPLAPNDNTSLYDCCINSLLQYFQNNRTLLLTANQ-----L 509
Db 456 -RTITLDVCCYRPMGCG---YDCLIMSPTNFOGKEHLDMKSNKEETVSDBDADYF 510
QY 510 MGQTSQVMDKDHFLVANAPLTFKQGTALALSCMDYGAVPVFPPLAIGYKGDYGEAEA 569
Db 511 SSEATTDENMNHAACTDQPMQOK--TKSGLSCMGTYGSPAPNM-VFGKNSINHOQANS 567
QY 570 LINTFSLANNYPAGDPELAQAKLMEBAFLBEMBAFORRMAGMPQVTFARSLDEINRTT 629
Db 568 IMMTILVYQ--RTEBPIQKALMEKEFLKCEYREKSPKVI-FSPMAERSTIDEIENDA 624
QY 630 AEDLPFATSYVIFLYISLALASY-----SSMSRVMDSKATLGLGVAVVLGAVMAAM 684
Db 625 KDEIVTVVIALAFLIGYVTFSLGRYFVCENQLMS--ILVHSRICLGMYSIINLSGFCSM 683
QY 685 GPFSTYIGRSSVILVQVVPFLVLSGADNIFIFVLEY--QRLP---RRGEPREVIHIGA 739
Db 684 GIFSMGFIHPVKNALVVOFPFVVTLLGVCTFEMVVKYVQAQRVSMYMSDQCEI-VGAV 742
QY 740 LGRVAPSMLLGSLSEALCEFLGALFPMAPVTFPALTSGAVLILDLQSAFAVALSLDS 799
Db 743 MAGTMPAMSSSLGCAFSPFIFGFTDPLAIRFCLYAGLAVLIDVLTHTCTILALVMT 802
QY 800 KQGEASRLDCCVCKPQELPPEGQEGE---LLG-----FQKAYAPFLHMI 844
Db 803 QRELNG-----KP-EFFFPYQIKDLGAYLIGORATDFEMTQFFHFQVAPFLMHM 853
QY 845 TRGVVLLFLALFGVSLYSMKHISVGLDDELALPKDSYLLDFPLNRFEBVGAARYFT 904
Db 854 TRITGIPIAFITTVIILSKISVGFDSMAFTEKSYISTHFRYLDKFFDGPAPFV 913
QY 905 TLGVNFSBAGNNAJCSAGCNNSFTOKIOYATEPPEOSYLAIIPASSWVDDPFIIDLTP 963
Db 914 DGEIDMHRDVOVKKFTFPFGCDPSTSGNIMNTAVAGHTEQTYLSGEMYMNIIDYLEISRK 973
QY 964 SSGCRLYISGPNKDFCPSTVNSL-----NCLKNCM-----SITNGSV---RPSVEQPH 1009
Db 974 SPCKKYVADPN--TFCSINRNKMSALDKACRTCMDPVDVANSYPKSINMYHRPSIEVFY 1031
QY 1010 KYLPWFLNDRPNKICPKGGLAAYSTSVNLTSQGVLASFPMAVHKL--KNSQDYTEALR 1067
Db 1032 RHLRHLLEDTPNSECEVGRASFKOAIPTSGRIGQASOFMFTHKKLSISNSDFTKAND 1091
QY 1068 AARELANITADLRKVPGRDPAEVPPPTITNVFEQYLTILPEGLFMLSCLVTPFAVS 1127
Db 1092 TABMWSRLERST-----DTRAHVFAYSKIFPEFQYITIMPLITQQLFTYVGVFGII 1145
QY 1128 CLLGLDLSGLLNLSIWMILVDTVGFMAWDISYNAVSLNLSAVAGMSVEFVSHIR 1187
Db 1146 CVTLGIDVGAACAVICQVSNYFHVIAFVYFINIPNALSATNVLVMSGILIEFSVNLK 1205
QY 1188 SFAISTKPTWLERAKATISMGSAVAGVAMTNLPQILVLGLAKQQLQIFFRNLMLLT 1247
Db 1206 GYACSLRQAKDRABETVSGIPILISGVVVMAGSTMPLSGAHQIITVYFPKFLIYI 1265
QY 1248 LGLHGLVFLPVIIISYVGPVDVNPALALEQKAEAVVAVVAWVASCNPHRSRTADNITV 1307
Db 1266 VSSAVHALIILPILAFGGSRGSGSETSTNDNDEHDACVLS--PTASHISNVBEGLI 1323
QY 1308 N-----HSFEGSIKGAISNPLPNNGRPF 1332

Db 1324 NRPSLLDASHIIDPLKAEIGIDKAT---GRDF 1353

RESULT 4

S52525

probable membrane protein YPL006w - Yeast (Saccharomyces cerevisiae)

N/Alternate names: hypothetical protein LPA11w; hypothetical protein YP03132.07

C/Species: Saccharomyces cerevisiae

C/Date: 08-May-1995 #sequence_revision 21-Jul-1995 #text_change 09-Jul-2004

C/Accession: S52525, S59687

R/Baddock, K.; Churcher, C.

submitted to the EMBL Data Library, February 1995

A/Reference number: S52519

A/Accession: S52525

A/Molecule type: DNA

A/Reads: 1-1170 <BAD>

A/Cross-references: UNIPROT:Q12200; UNIPARC:UPI000006A19C; EMBL:Z48483; NID:g683777; PID:

R/Experimental source: strain AB972

R/Hall, J.; Ahmed, A.; Bussey, H.; Fortin, N.; Friesen, J.D.; Storms, R.K.; Vo, D.H.; Wa

submitted to the EMBL Data Library, August 1995

A/Description: The sequence of Saccharomyces cerevisiae chromosome XVI left arm.

A/Reference number: S59677

A/Accession: S59687

A/Molecule type: DNA

A/Cross-references: UNIPARC:UPI000006A19C; EMBL:U33335; NID:g965076; PID:g965087; MIPS:Y

C/Genetics:

A/Gene: SGD:NCRI

A/Cross-references: SGD:S0005927; MIPS:YPL006w

A/Map position: 16L

C/Keywords: transmembrane protein

F.1-17/Domain: transmembrane #status predicted <TM1>

F.258-274/Domain: transmembrane #status predicted <TM2>

F.341-357/Domain: transmembrane #status predicted <TM3>

F.585-601/Domain: transmembrane #status predicted <TM4>

F.614-630/Domain: transmembrane #status predicted <TM5>

F.667-683/Domain: transmembrane #status predicted <TM6>

F.698-714/Domain: transmembrane #status predicted <TM7>

F.751-767/Domain: transmembrane #status predicted <TM8>

F.1004-1020/Domain: transmembrane #status predicted <TM9>

F.1027-1043/Domain: transmembrane #status predicted <TM10>

F.1051-1067/Domain: transmembrane #status predicted <TM11>

F.1103-1119/Domain: transmembrane #status predicted <TM12>

F.1137-1153/Domain: transmembrane #status predicted <TM13>

Query Match 19.2%; Score 1329; DB 2; Length 1170;

Best Local Similarity 28.8%; Pred. No. 1,1e-84;

Matches 373; Conservative 230; Mismatches 512; Indels 182; Gaps 46;

QY 10 LLMAL-----LTLAASEPYTTIHQPGYCAFYEDEGKNP----ELSGSLMTLSNVSCLSN 60

DB 3 VMTIITLVGQIMLVLQG-----TATCAATGNGCKKSVRGNELPCPFRSPBPVLSLD 54

QY 61 TPARKITGDHLLLOKICPRLYTGPNTQACCSAKOLVSLKASLITTKALLTRCPACSDNF 120

DB 55 ERSK-----LLVEVCGSEWKEVR--YACCTKQGVVALRDVLQKAQPLISSCPACKLKNF 105

QY 121 VNLHCNHTGSPNOSLPFNTNTRVLAQAGQLPAVVAYEFYQHSFPAQSDSCSRVRVPA 180

DB 106 NNLFCHFTCAADGRFVNITKV--EKSKEKDKIIVELDVFNMSWASEFVDSCKNKKFSAT 164

QY 181 ATLAVTGMGVYSALNNAORWLNFGQDTGN--GLAPLDITFHLLEPGAVSGSIQPLNE 238

DB 165 N-----GYADDLIGGAQNYGQFLKFLGDAKMLGSSPQINIKYDLANEE--KEWQEPND 218

QY 239 GVARNESQGDVATVATCSQDCAASCPAIARQALDSTFYLQMPG---SLVLIITLCSVF 295

DB 219 EYVACDAQ----YKCAQSCDQESCPHL-KP-LKDGVCKVGLPFCFSISLVLIIFYTICALF 272

QY 296 AVVTILLVGRVAPAPRKSKVNDPKGTSLSDK-----LSTRTH-----LLGQFFQCGNG 345

DB 273 AFPMVYILCKRKKGAMLVDDDIVESG--SLDESETNVESFNNEENTFNGKLANLEFTKVG 331

QY	346	TMVASMPLTITLVSVIPVVALAAALV-FTELTDPVLMASPNQASSEKAFHOHCEPF	404
Db	332	QPSVENPKLITLVTSIPVFSFIIFQYALTEITPILNLMVSKNEKEKEKFFPDNKEP	391
QY	405	FRTVQVILITAPNRSRYRDSL--LLGKPNFGSGLIDLLELLELLEOERLRLQVWSPAQ	462
Db	392	YRTQIQIPVNVETGVLSSETLHMVFENF-----ITBEL-----QSS	429
QY	463	RNLSLODI CVAPLPNDNTSLYDCCINSLLQYFQNNRLLLLITANQTMGOTSQVDMKOHF	522
Db	430	ENIGYODICFRP-TEDSF---CVIESFTQYFGQ-----ALPNKMSWREL	470
QY	523	LYCANAPLPTFDGRTALMSCADGAPVFPFLAIGYKKQDYSEBALIMFSLNNYPAG	582
Db	471	QECGRP-----VNCPLPTQDPKTNL---PSDDILNAHAFVTLITLNH---	514
QY	583	DPRLAQAALMEAEALTEEMRAFORHMAQFQYFTFAERSLEDEINRTTAEDLPFATSYIV	642
Db	515	---TQSANRMEER-LTEYLIDLKVPBGL-RISFNTEISLEKELNN--NDISTVAISLM	567
QY	643	IFLYISLALSGYSMSRWVNDKSKTLGLIGVAVVLGAVMAAMGFSYIGIRSSLVILQV	702
Db	568	MFVATVATLRRKDKTRUL-----LGISGLIYASIVCAAGELTFLGLSTIIIEVI	621
QY	703	PELVISGADNIFIFVLEYOR-LPRRDPREPHVIGRALGVASMSLCSISEAICFPLG	761
Db	622	PFLIATGIDNIFILTHEYDRNCQKPEYSIDQKIIAISGMSISILMSLCLQCGCFIA	681
QY	762	ALTEMPAVRTFALTSGLAVIDLFLQMSAFVALLSLDSKQBSRLDVCCKVQDELPP	821
Db	682	AFVMPAVAHNPAIVSTVSIENGVLQTLAYVSIILEKSNYKQIT-----	728
QY	822	GQSGSLILGFQKXKAPFLHMITRGVLLFLFGVSLYSMCHISVGLDQELAPRDS	881
Db	729	-GNEETKESFLKTYFKMLTQ--KRILIIIFSAMFTSLVFLBEIOFGIDQTLAVPDS	784
QY	882	YLLDYFLFLNLYFEVGAIVYFVTTLGNFSEAGNMAICSS-AGCNNFSFTQKIQVATEF	940
Db	785	YLVDFKDYVSFLNAGPVVYVVK-NIDLTRQNOQKICGFTTCERDSLANYLE---QE	840
QY	941	PEQSYLAIPASSWVDFIDMLTPSS--CCRLYISGPNKDKCEPSTVNSLNCXKMSITM	998
Db	841	RHRSTITEPLANMLDDYFMFLNPQDCCRL--KKGTDEVCSPSPSRRC-ETCFQ--Q	894
QY	999	GSVR-----PSVQPFKYLPMFLNDRPNIKCPKGLAYSTVNLSDQVLASRFAY	1052
Db	895	GSWVYNSGFEEDGFMEYLSIWIN-APSDCPLGGRAVPSTAL-VYNETSVSASFETA	952
QY	1053	HKPKNSODYDEALRAARELANITADLRKPGTDPAFVEPPYITVNFYEQYTLIPEG	1112
Db	953	HHPLRSQDFIQAY-----SDGVRISSSFPELMDMAYSPYIFFYQYQTLGFLT	1001
QY	1113	LFMLSLCLVPTFAVSCLLGLDRLSGLINLISYMLVDYVGNALMDISYNAVLNLV	1172
Db	1002	LKLIGSAIILIFFISSVFL-QNIRSSFLLALVYMIITVDIGALMALGISINAAVLNLI	1060
QY	1173	SAVGSVSEFVSHITRSFAI---SRKPTMLEBAKEATISMGSAVAGVAMTMLPGIVLGL	1229
Db	1061	ICVGGVGEFCVHIVRSFTVPSETRKDANSVLYSLNTIGSVIKGILTKFIDGVCLAF	1120
QY	1230	AKAQLIQIFFPRNLILITLLGLLGLVFLPVILSYVG	1266
Db	1121	AQSKIPDVYFRMWFTLITVALLALLFLPLALSLFPG	1157


```

Db      765 IACYCTLCLEFGNOMABSEVCITFSHANGSFQSYLTFRQNNENINIGPFLMFW 824
      905 TLGYNFSSEAGMNAICSSAGCNNSFTOKIQ---YATEPEOSYLAIPASSWDDFLDWL 961
      825 EGVKXKHPDKQKQKFCCTLAGCDNSMGNKIRSLAYAMNY-KGNVYHQDVIMWDSYQFM 883
Qy      962 TP-SSCCRLYISGPNKDKFC 980
      884 HPRGSCCKM-----DGKQFC 898

```

RESULT 8

```

T25600
hypothetical protein C32E8.8 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T25600
R:Gating, S.
Submitted to the EMBL Data Library, February 1997
A:Description: The sequence of C. elegans cosmid C32E8.
A:Reference number: Z20056
A:Accession: T25600
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-933 <GAT>
A:Cross-references: UNIPROT:P91129; UNIPARC:UPI0000611CF; EMBL:U86308; PIDN:AA842325.1;
C:Experimental source: strain Bristol N2; clone C32E8
C:Genetics:
A:Gene: C32E8.8
A:Map position: 1
A:Insertions: 145/1; 177/2; 385/1; 418/3; 833/3; 878/3

```

```

Query Match      8.7%; Score 598.5; DB 2; Length 933;
Best Local Similarity 22.9%; Pred. No. 1.2e-33;
Matches 251; Conservative 181; Mismatches 362; Indels 303; Gaps 37;

Qy      341 FQGGTGVASWPLTIVLSVIVPVALAGVFTLTDPELWASQARSEK-AFH-- 397
      14 FROLGFLICDHPFPFPPFLFTAMGVLHLPDSDAVLYFTPLGQSMERMSIHEK 73
      398 ----DOHFGP---PFRTNQVILFTAPNRSYRYDLSLGPKNFSGILDLLELLELQER 450
      74 WPLTDNNYIFGRNAVQSRHIOVTLARN---DSNIIDPKRANAVYQLDK-----IQTR 124
      451 LRHLQVMSPEAQRNISIQDICYAPLNDNTSLYDCINSLIQYQNNFTLLLTANQTLM 510
      125 VRVLT----- 129
      511 GQTSQVMDKHFLYCANAPLTPKQGTALA-----LSCMADYG---APVPPFLAIGY- 559
      130 -----NGHYGVKNLCQYKNGGCPSNKGVHILSDIHNHGFNITYPYFRFSGEGYI 181
      560 -----KKGKYS-----AEALIMTFSLNNP-----AGDPRLAQAQLEAFLEE 599
      182 GSSLGCVTVMKGENETDLASAKAFMYLHKFHEEMSYISGE-----WE---DEL 230
      600 MRAFORRMAGMF-QVTFPAERSLEDEINRTTAEDLPATPSYIVIFYLISLALGSY--SS 656
      231 GRMLTQYREDPYISTTHSQTLADELRANDTLIPRTIISTILLIVSTICSLSDIGS 290
      657 WSRVWVDSKATLGLGVAVVVLGAVMAAGFFSYIGIRSSLVILQVPPFLVSVGADNIFI 716
      291 FSIQWVLSKPLTSLISIGVVSAGAIITGVGLSLMGMPYN-DIVGVMPFLVAVGVDMFL 349
      717 FVLEVQRLPRRPGREVEH--IGRALGVAPSMILCSISSEALCPFLGALTTPMRAVTRAL 774
      350 MVAAN-----RTSRTHVHERMGECLADAANVSIITSSTDVLSFVGAIITTIIPAVQICV 405
      775 TSGLAVIDPFLQMSAFVALLSLDSKQESR-----LDVCCVCV----- 814
      406 YTGVAIFPAFIYQITTFPAACIALAMKHBASGRNLSFLIEAVSAEKKTSLSTFORLNLGS 465

```

```

Qy      815 -PQELPPQCGEGRLILGFPQKAYAPFLMHTRGVLLLF-LALFGVSLYSKCHISVGLD 872
      466 VPDHSAASHVQKQPLTSRFFGEMVAPPLMHPVVRGIMWVFVYLLGAS--YGSRIKEGLE 524
      873 QELALPKDSYLLDYFLNRYF-EVGAPYFV-----TTLGYNFSSE 913
      525 PVNLLVEDSYAIPHYLLERKYEMKYGOQVQIVINAPDLRNHTSRDRVHAMVLDFAFISKH 584
      914 A-GMAAICSSAGCNNSFTOKIOYATEPPEQ-----SYLAIPASSWDDFLDWL 961
      585 AIGMSV-----QFWLFEMERYQKELEVQIIDDSPFYGLHHFPLASKTNPLAEDLYW- 637
      962 TPSSCCRLYISGPNKDKFCSTVNSLCLKMSITMGVSPSEVQFHKYLFWFLNDRPN 1021
      638 -----GMPD----- 642
      1022 IKCPKGLAAVSTSVNLITSDQVLA-SRPAVYHKPKNSQDYTEALRAARELANITADL 1080
      643 -----DDNGTMVKSFRFLLGKQDLVTTMDQDADATMSPREVAARW----- 681
      1081 RKVPETDPAFEPVPTITNVFYEQYLTILPEGLFMLSCLVPTFPAVSCLLGLDRLSGLL 1140
      682 -----PEPNVTFPMPIWMFTDQYIIIPNTVQNIITIALVNIIVIAVFIPOPMCS-LW 733
      1141 NLSTVMILVDVYGFALMDISYNAVSLINLVASGVMSVEFVSHITRSFAISTKPTWLER 1200
      734 VALACASIDEGVIGYTWLGVNDALSMITITNISIGSVDSYSHIAYGVASSBETIAGR 793
      1201 AKERTISMGSVAPGAVAMTLPGLIVYGLAKAQLQIFPRLMLITLLGLLGLVFLPV 1260
      794 VKALSLAGWPLSQG-AMSTIIVASVLDADIPAIMI-VTFPKTVVLSISLGLHGLVFLPV 851
      1261 ILSYV-----GPDVNPALALEQKRAEA-----VAAVMVASCPNHPGRVS 1300
      852 LLSIFVRGCCIIIPSSPHGHSAQKIEQRIAMISSPDLRLVAPLRASSPSIFRRLR 911
      1301 -TADNIYVNSFEFSIK 1316
      912 YTDSPVTHNRKNSIK 928

```

RESULT 9

```

T13952
membrane protein ptch2 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C:Accession: T13952
R:Motoyama, J.; Takabatake, T.; Takeshima, K.; Hui, C.
Nature Genet. 18, 104-106, 1998
A:Title: Ptc2, a second mouse patched gene is co-expressed with Sonic hedgehog.
A:Reference number: Z17830; MUID:98122566; PMID:9462734
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1182 <MOT>
A:Cross-references: UNIPROT:Q35595; UNIPARC:UPI000001587; EMBL:AB010833; PIDN:BAA24691
A:Experimental source: strain BALB/c3T3c
C:Genetics:
A:Gene: ptch2
C:Superfamily: Drosophila membrane protein patched
C:Keywords: transmembrane protein

```

```

Query Match      8.4%; Score 583.5; DB 2; Length 1182;
Best Local Similarity 24.2%; Pred. No. 1.8e-33;
Matches 282; Conservative 154; Mismatches 433; Indels 297; Gaps 44;

Qy      330 SFSHTILGQ-----PFGG-----WGTWVASWPLTIVLSVIVPVALAGVFTBELT 376
      20 SSAPHIILAGSLQAPLWRAVYFQGLFSLGCRIQHCGKVLFLGLVAGALLGLRAVAVIE 79
      377 TDPVLMASPNQASRSEKAFHDQHG-PFRTNQVILTPAPNRSYRYDLSLGPKNFSGI 435
      80 TDLRQLVVEGSRVSGELHTYKELGEBAAITSGMLT-----QTHAGSGNVLTPE-----A 131

```



```

QY 996 ITMSVRSVBQFHXYLWFLNDRNICKCPKGGIAAYSTVNLSDQGV--ASRFMAVHK 1054
Db 630 -----VFLSQPQF-----AKNRPDVLLTJEDGHELEASRI----- 658
QY 1055 PLK-----NSDYTEALRRARELAANTADLRKRPCTDPAFEPFPYTTINVEEQYLTLP 1110
Db 659 PVQLRHVSSAANSQSRMKLPRLAE--TSELQ----TGYYADFQ-----PAEQYNAVLP 706
QY 1111 BGLFMLSCLVPTFAVSCLLGLDLRSGLNMLSIYMLVDTVGFMLMDISYNAVSLIN 1170
Db 707 GTLSSIAAGVAVVAVLSLILPEVAS-LWVSFSLVSNIGILGPMFWSVRDLFISMVT 765
QY 1171 LVSAVGMSVEFVSHITRSFAISTKPTWLERAKKATISMGSAVFAVQANTNLPGLIVGLA 1230
Db 766 IYMSIGFCVDPFAHLAYNFAKGQNMDSERBMNLYAVGADPLMS-ATSTIIIGVSFMA5A 824
QY 1231 KAQLIQIFFRRLNLTITLGLHGLVFLPYLS--YVCPDVNPALALEQKAAEAVNAV- 1287
Db 825 ESYVFR-SFLTIMIVILLGALHGLVILPVLLSMFYCGG-----SSKKAKEHDAVD 875
QY 1288 -MVASCPNHPRSVST 1301
Db 876 QKLOQOYNNPAPTAS 890

```

```

RESULT 11
T18291
patched protein - zebra fish
C1:Species: Brachydanio rerio (zebra fish)
C1:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C1:Accession: T18291
R1:Concordet, J. P.; Lewis, K. E.; Moore, J. W.; Goodrich, L. V.; Johnson, R. L.; Scott, M. P.;
Development 122, 2835-2846, 1996
A1>Title: Spectral regulation of a zebrafish patched homologue reflects the roles of sonic
A1:Reference number: Z18860; MUID:96379744; PMID:8787757
A1:Accession: T18291
A1:Status: preliminary; translated from GB/EMBL/DBJ
A1:Molecule type: mRNA
A1:Residues: 1-1220 <CON>
A1:Cross-references: UNIPROT:Q98864; UNIPARC:UPI00001328B3; EMBL:X98883; PIDD:CAA67386.1
C1:Gene: ptc1
C1:Genetics:
C1:Superfamily: Drosophila membrane protein patched

Query Match      8 3%; Score 572; DB 2; Length 1220;
Match Local Similarity 22.4%; Pred. No. 1.2e-31;
Matches 270; Conservative 174; Mismatches 431; Indels 328; Gaps 43;

QY 321 KGTSLDRLSTHTLLGQFPGWGTVASWPLTTLVTSVPAVALAGLFTLETDPV 380
DB 51 KGVAVGQAPLPIRRPQAFLESLGCHIQIRCGKVLFTGLLVGALSTGLRVALETIDIE 110
QY 381 ELWSAPNSQASSEKAFHDQHFQ-PPFRITQVILTAPNSSSYRYDSLGLGPNFSGILLDD 439
DB 111 KLMTWAGSRVSKELRYTKKEQGEBSVFTSQMLIQPKQEG-----TNILTQE 157
QY 440 LLLLELLEQERLRIHQVWSPFAQRNITSODICTAPLNF-DNT-----SLYCCINSL 490
DB 158 AL-LHLLEALASAKQVNSLYGKSWMDLNKICFQSGVPIIENWMIERNIDLFPCCMITVP 215
QY 491 LQYFQNNRTLL-----LLLTANQ--TLWGQTSQVDMKHPIFYCA 526
DB 216 LDCFFEGSKLQGGSAIYLPQMPDIOQMNIDPLKLEHLSQPTSLSGFREMDDKAQVGHAYM 275
QY 527 NAPLTFKDGITALAISCMDYGAQ-----VPPFLAI--GSYKG----- 561
DB 276 NRPCLDPSDT-----DC-PSHAPKNDPQOVENIAAELQGGCHGFSKKFMHQEELILGBR 329
QY 562 -KD-----YSEAALIMTFSL-----NNVYAGDPRLAQAQKL-----WEAPLLE 599
DB 330 VKDSQNALQSAEALQTMFLMSPKQLYEHFKYDQYIHIINNEDKATLAILESQKQKFEV 389
QY 600 MRAF--QGRMAGMPQVTFATERSLDEINRTTAEDLPFAFNSYIVIFLYTSLAAGSYSSW 657

```

Db	330	VHGSI	PQWSSSNVAFSTT---	TLNDIMKSRSDSVIRVAGVILMLALYACVTM---	LRW	443
Qy	658	SRVWVDS	KATVIGLGAVAVVIGAVNAA	MGFSYIGIRSSLVILQVVPVLVLSVGADNIFIF		717
Db	444	D--CAKSG	AGVAGVILVLTALSVAA	GCISLIGLSFNAATTQVLPISALGIDVDMFL-		500
Qy	718	VLEVQRL	PRRGRGE	RVENHIGALRVARVPSMLGCSLEAICPFGALPMPVPRFALTSG		777
Db	501	-LGHSF	ETERNINIPKERTGDC	LRRGTGTSVALTISVNMMLAFPMALVPIPLRAFSLOAA		559
Qy	778	LAVILDF	LQASAFVALTSLDSKQ	QEASRLDVCC-----VKPOE-----		817
Db	550	VVVVFN	PMALLIPALISLDI	HREDKRLDILCCFYSPCSSRIYQIQPQELSDANDHQ		619
Qy	818	---	---	---	---	823
Db	620	RAPATPY	TGSTITTTSTHITTTVO	AFQCDAAQHIVLIPRQISITTPPSMWLSTPTP		679
Qy	824	---	---	---	---	846
Db	660	TTDPYGSQ	VFTTSSSTRDLA	QVEBPKEGRCVLPFFRNMWLSFAEKAPALLKRETK		739
Qy	847	GVLILL	FLPALFGVLSYMSCHISV	ELDOBLAPKDSYLDYFLPNRRYFEGVAPYFVYTL		906
Db	740	TVVVVVF	PALLSLSTLXGTMW	HDLYLTDIVPRTOGEYERITTAQFKFSF-YNNYLVITMD		798
Qy	907	GYNFSSE	AGMNAICSSAGC	NNFSFTQKIQVATEFPEOSYLA-----IPASSWDDFLDW		960
Db	799	GFDYAR	---	---	---	841
Qy	961	LTPSSCC	---	---	---	993
Db	842	LKGLQAT	FADADWEAGKITT	YSYRNGTEDEGALAYRPLIQTSKKEPFYSSQLTSSKRLVDG		900
Qy	994	MSITMG	SVRPEVOFHKYLP-W	FLNDRPNIKCPKGGALA-----YST---		1034
Db	901	---	DGLIRPEV--FYI	YLVTVNSD-----PLGYAASQANFYPHRREWHDXDPTTGE		948
Qy	1035	SVNLTS	DGQVLASRFMAVHKP	LKNSQDYTEALRARELAAN-ITADLRKYVGTDPABEVF		1093
Db	949	NLRIPAR	BPLLEFAQFPFYL	NGLRQASDFIEAIBSVRTICEFHMQGIKNYPNG-----Y		1002
Qy	1094	PYITTNV	YEBQYUILLBEG	LFMLSLCVVPFAVSCILLGDRSGILLNLSIVMLVDTV		1153
Db	1003	PF---	LFWEYDIGRHW	FLISTISIVLACTFLVCAILLNLPMTAGVI-VFILLPMYVELF		1057
Qy	1154	GFMALMD	SVYAASVLINLSA	VGMVSVEFSVHITRSF--AISTKPTWEAKAERTISWGSA		1211
Db	1058	GIMGLIG	KLSAIFRIVILIAS	VGIGVEFYTHALAGFLTALIDRNT---RSANVMBEHNAP		1114
Qy	1212	VFAGVAM	TNIPGILVLGLAKA	QOLIQIFEPRLNLTITLLGLIHGLVFLPVILSYGP--DV		1269
Db	1115	VIDG-A	ISTLLGVMLGSEB	DFIMRYFAVAILTLLGLIGLGLVLLPVILSLMGPARAV		1173
Qy	1270	NPA	1272			
Db	1174	VPA	1176			

RESULT 12
T29550
hypothetical protein F55F8.1 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T29550
R:Gatlung, S.; Scheet, P.; Kemp, K.
submitted to the EMBL Data Library, November 1996
A:Description: The sequence of *C. elegans* cosmid F55F8.
A:Reference number: Z20647
A:Accession: T29550
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA

A:Residues: 1-889 <GAT>
A:Cross-references: UNIPROT:P91346, UNIPARC:UPI000017BA20, EMBL:U80447, PIDN:AA837812.1,
A:Experimental source: strain Bristol N2; clone F55F8
C:Genetics:
A:Gene: CESP:F55F8.1
A:Map position: 1
A:introns: 36/3; 66/3; 98/3; 149/2; 200/3; 240/2; 279/1; 358/1; 481/2; 569/3; 606/3; 697

Query Match 7.8%; Score 539.5; DB 2; Length 889;
Best Local Similarity 23.0%; Pred. No. 1.5e-29;
Matches 223; Conservative 156; Mismatches 337; Indels 253; Gaps 35;

QY 401 FPPFRPTNYYL-----TAPNRSSYRDSLLGPKXPSGLDLDLLELE 446
DB 31 FGPYSYTERRIIDAMPVLDGTFVAGRAVTOG--REVQVAVAARSGNLDKRVSNELKL 88
QY 447 LGERLRHLQVMSPEAQRNLSLODICVAPLPNDTSLDYDCINSLLQFQNNRTLLLTAN 506
DB 89 MSFTRN-NITVOFSKRTWSFADLCLAA--GPDGR---CANNDDHIO----- 127
QY 507 QTLMGQTSQVDWKDHFELYCANAPLTFKDGDTALALSCMADYGAPVPPFLAIGYK----- 560
DB 128 --LASRLHQ-----HGINITYPTVRLSKSAVLAS-----ALGGVKLAKGDN 167
QY 561 GMD-YSEARALIMTSLNLYP-----AGDPRLAQAKLMEZAPLEBMRAPFORRMAGMFOY 613
DB 168 GENIIVEATAMLLIYQLKFPNEISVSG-----LMEREKKNMDEY-KKQAKYTSI 218
QY 614 TPTAERSLDEINRTAEDLPFATSYIYIFLY-----ISLALGYSWSMRVDSKAT 667
DB 219 TYFHSQTLSDLEINRNAERLAPKFIGAFVILVCPVSILCTYTKGSGYIDM---VTKFI 274
QY 668 LGLGVAVVLGAVMAAMGFSSYLGRSSLVLIQVVPFLVSGADNIFIVLEYOQLRPR 727
DB 275 LSVLGSNNGMGIASMGMLTYLEIQYN-DITAVMFLVAVGDTMFLVMSLKRTDR- 332
QY 728 PEEPRERVHIGRALGRVAPMLCSLEALCFELGALTPEAPVATPMTSLGLAVILDFLIQ 787
DB 333 -MLKYDORLACMADAVALIITALTALDSFGVGITITIPAVQIFCIYMCALLLTFAYO 391
QY 788 MSAPVALSLDSKROE-----ASRLDY-----CCCVKROELPPGOG----- 824
DB 392 LTFPCALIVYYTRIEQGLHSIWLRAVYTSSTSPNLVLFMLGSGOPQK-PLPSCCTVSS 450
QY 825 -----EGLLGFPQKAYAPFLH--WITRGVLLFLALFGVSLYS 863
DB 451 TSSVSTMTSQATSPASKLHHCAGATSPFRNMYAPVLMQPMI-RAIAGLVLYLYLGISTIG 509
QY 864 MGHISVGLDOELALPKDSTLLDYFLFLAN-YFVGVAPVYFVTT-----LGYFSE 913
DB 510 CTHLEKGLERPANLLVDDSYATPHTVLEKGYWHTYGASLQIVVSNPDLRPVERIMDMK 569
QY 914 AGMNATICSSA-GGNNSF-----TQKIQYATE-----PPEOSYLAIPASSWVDFI 958
DB 570 ASTPANCKVAILGDDSYQFWLREMOUSEEIHKIQYDNEKEYDHAQYIYSWDMGQPMVADV 629
QY 959 DWLTSSCCRLYISGENKDKFCPSTVNSINCLKNCSITMGSVPSVEOFHKHKLWFLND 1018
DB 630 -----WGRKN 634
QY 1019 RPNICKPKGGLAAYSTSVNLTSDGQVLASRFMAHYHKKPLKNSODYTALAAARELANITA 1078
DB 635 -----NSBRILIKTRPFMIGMDISTTTKQTATATTFREIASRF-- 672
QY 1079 DLKVPYGTDPAFVPPYITTNVFEYOXLTLPEGLFMLSCLVPTPAVSCLLGLDLSRG 1138
DB 673 -----EQYNVTTYMPLMLFTDOYALVENVMTODIIVAAQMLVISALLIPQVCSF 723
QY 1139 LINTLSIWIWLVDTGVFMALMDISYNAVSLINLVASVMSVEFVSHITSPAISTPMTL 1198
DB 724 WY-AVYIGSIDGVLGFMFLMNVNDAISMITTIMSVGSVDYSAHITAYVYSKSTTS 782
QY 1199 ERAKEATISMSGAVFAGVAMTNLPGILVIGLAKAQLIQIFPRLNLLITLLGLHGLVFL 1258

DB 783 ARVCALDLEMPVAQG-AMSTILAVSVLSDVPAYMI-VTFPKTVFLAISIGFLHGLVFL 840
QY 1259 PVLIS-VYG 1266
DB 841 PLMLSVFVG 849

RESULT 13
T27969
hypothetical protein ZK675.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T27969
R:Sim8, M.
submitted to the EMBL Data Library, November 1994
A:Reference number: Z20448
A:Accession: T27969
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1405 <WLL>
A:Cross-references: UNIPROT:Q09614, UNIPARC:UPI00000835A5, EMBL:246812; PIDN:CAA86843.1
A:Experimental source: clone ZK675
C:Genetics:
A:Gene: CESP:ZK675.1
A:Map position: 2
A:introns: 31/1; 59/3; 100/3; 244/2; 485/1; 950/2; 1298/2; 1335/3
C:Superfamily: Drosophila membrane protein patched

Query Match 7.4%; Score 514; DB 2; Length 1405;
Best Local Similarity 20.2%; Pred. No. 1.7e-27;
Matches 267; Conservative 190; Mismatches 412; Indels 452; Gaps 52;

QY 331 KQTSLSRLSSTHTLLGQFFQMGWVY--ASWPLTILVLSYIPVALAAGLVPTLTTD 378
DB 100 RQKATGNRYALYSRLQLKLFALGNVTHRRAMS-ILAVMIFAVC-CYGLQYHIETD 157
QY 379 PVELMSA-----PNSQARSEKAFHDQHPFP-RTN-----QVITAPNRSS 419
DB 158 IYKLVAAQGRLEDEINLFLPNIKEMARNVTGDS--GPELPRENGLGSGGYQLIQPEYEG 215
QY 420 YRYDSLGLPKNFSGIIDLLELLELLELRLHQVSPKQRY-----SLQDICY 472
DB 216 --QDLAAGP-----LKKHVEIKHITASFVNSVHGVDWMSLDICF 253
QY 473 APLNDP-----NTSLVD-----CC-INSLQYFQNNRTL-----LTL 503
DB 254 KPAPSVAAADSAASSLGVIDKIYPCIWITPIDCWEGSKALGPHPSLPKSLGPIGLML 313
QY 504 TANQTLMGQTSQVDWKD-----HFLYCANAPLTF----- 532
DB 314 SS-----LSDGMIRWSDPDIAVIDEIRSFNLGSHYFFPERAGVSHGYMDRPCIIDLP 369
QY 533 -----KDGTAALSCMADYGAPVFPFLAIGYKGYKQYSBAE 568
DB 370 BCPMAKYPDVCPIHDVREIAKKYGTLEBEKKDGSYFPDPL---GRKGRAGQ 426
QY 569 ALI----- 571
DB 427 KMIHPAQPADSIPTIEDAVPAQVPVSTAPIPTTTTSLSEBARAAEKEKKQKARELKYC 486
QY 572 -----MTFSLNNYP-----ADP 584
DB 487 KSYRKAPEWMLKKNDKKPEVWSENNYQNVDAEMTGCGSFASVNLNPEDMILNIP 546
QY 585 RLAAQ--ATL-----WEEAPLEE-MRA 602
DB 547 BRACKGKLSGADALQSVFLVASPADVFLRFKQKGRNSMKTKGLMDAMNETAAEQVLOA 606
QY 603 FORNAAGMFOVYTFARSL-----DEINRTAEDLPFATS-----YIV 642
DB 607 WQR-----NFT--KSLYNHKKANVDEDDGNERRLTH--PLASTISIDMLBERFCQFNNTI 654

[illegible]

Query Match	6.9%	Score 473.5	DB 2	Length 1015
Best Local Similarity	19.24	Pred. No. 7.5e-25		
Matches 226	Conservative 186	Medium 417	Indels 347	Gaps 38
QY	289	IIILGSVAV---VYIIILVGFVVAAPARAKSKMVDPKGKTSLSDKLSFSTHTLLGQFQGWG	345	
Db	11	LLKTSIFBAGKNVLLALIAHRF--PMRCRIPLDCK---LEKVAWYTKHLVDY----	61	
QY	346	TVVAVSWPLTLLIVSLVFPVVALAAGLV--TEILT-----DPV-----ELMSA	385	
Db	62	-----YVPLIAVILITAIKSCGFWIKETLLDARKLTPVSAIPMKEQVSELMPV	114	
QY	386	PNSQARSEKAHQDHFGPFRTNQVILLTAPRSSYRDSLLGPKNSGILDDLLEL	445	
Db	115	KSYEFLEPRT-----FQMNRYLIVVHGREGNGT-----YPNILEGSYLANDIA	158	
QY	446	ELQERL-----RHQOWSPBAQRNLSLQDICTAPLPNPTSLYD	484	
Db	159	KIEBELVTNVSFPREALEDNPALMRKRHLNETEIEENRNTI-----TFKD	205	
QY	485	CCINSLLQYPONNRTLLLTANQTLMOGTQVPMKHPIVCANAPLTPKOSTALALSMA	544	
Db	206	VCMWVYGDCKREKNVIBLKRRLHN-----KGISVTPQVN	243	
QY	545	DYGAVPPEPLAIG---YKGDVSEAEALIMTFSLANNYPAGDPRLAQAQKME---EAPLE	598	
Db	244	QEGPIYIAFYIGVDTYENDTITKYAAMRLWYFLKDDDEQEOU--AKFEDIAEKTVR	301	
QY	599	EMRAPORRMAGMPQVTFPTEARSLDEINRTABDLPITFATSUYIPIFYISLALSYS---	655	
Db	302	ETVADHPIT---QCHKHSRIYDQGLTRANNRKLKPFNVNTIANLVLEFAN---YSKWK	353	
QY	656	-----SWSRWVD--SKATLGLGVAVVLGAVMAAGFFSYGIRSSVLIIOVPPV	706	
Db	354	YFRMDHWP--LHIDWLSKRPVLAAGVLSVIALISGILLWPGMEFAEITL--IAPFLV	411	
QY	707	LSVQADNIFIVLEY----QRLEPRRPGEPREVIHGRALGRVAPSMILCSLEAICFLPGA	762	
Db	412	LSIGVDDMFIIAAMHNTKMPKPSRSPKVMKRMIIEMASESAVAIPIFSPFDVLSFGAGT	471	
QY	763	LTPMPAVRTPLTSGLAVIIDFLLOMSAFVALLSDBKROQASLNDVCCV-----	813	
Db	472	ITDIIAVOGFCAMTAAOCMFTFELYQITFPALMVISAKAQMSSGRNCSMPCITAGDIYIE	531	
QY	814	-----KPOELRP	820	
Db	532	DGSLQPNLKKKKSKRKDAKAEKKEDBAKDKMEIBENNAENAEKSYDSSPPSQIHI	591	
QY	821	PGQSEGILLGFQKAVAPFLHMTITRGVLLFLFALFGVSLYSMCHISVGLDOBELAPKD	880	
Db	592	PVKSRGAMGHFRDIYVPMILNMTKLFPMFTPIIYIGISVYIGICWNEGQSDYDKLLHS	651	
QY	881	SYLIDYELFLNRYEYVGAPVYFVTTLGYNFSSAQMNAICSSAGCNNSFPQKIQYATEF	940	
Db	652	DPIVE---ALGRREIL-----FHGGDQLEIA---FHGGDQLEIA---	674	
QY	941	PGQSYLAIASSWDDFDIMLTPSSCCRLYISGNPKDPCPSTYNSLNLCKLCKMSITMG	1000	
Db	675	-----IONCENTITAE	685	
QY	1001	VRPSEVQFHKLPMFLNDRPIKPKG-----LAAYSIVNLTSD-----GQ	1043	
Db	686	SRDRIE-----LVAQFENISYSLGKGTSFWMREKXKYSNLTGSYLANDRESWITGV	738	
QY	1044	VLASFPMAVHK-----PLKSGOY--TEALPAAR-----ELAAVITADLRKVP	1084	
Db	739	YEMSOLEAFYGLMSODFYMANESDYDLELSYFRIGVNRHLSIPTDLVL--ITSELKGYA	797	
QY	1085	GTDAFEVFPRTTINVFYEQYLLIPBGLFMLSICLVTPFAVSCILLGLDLRSGLNLS	1144	
Db	798	DRHEDLNIIVYQOSRAILDQNLVLLSITNTMDTLAMCMFCVALLIPIPNP--CALFITFA	856	
QY	1145	IWMILVDTVGFMAIWDISYNAVSLINIVSAVGMSVEFVSHITRSFAIS--TKPTMLERAKE	1203	

Db 857 WTITDGVGFSLWSVKLDPISMTITISGSIIFSASHITHGFSNDSNLSAFRCVD 916
Qy 1204 ATISMGSAAVAGVAMTNLPGLIVLGLAKAQLIQIPEFRNLITLGLHGLVLEPILLS 1263
Db 917 AMEKLAMPVHVG-SLSTIGGVFLARIDSYVMVLVFETKISIVL-IIGAHMLMLPILLS 974
Qy 1264 YVGP-----DVNPALALFQKRAEEVAAMVASCP 1293
Db 975 MCPIVTERLSDASKKASDRRRKLEKNKNSVYAINLP 1010
RESULT 15
S06119
membrane protein patched - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: S06119; A33468
R:Nakano, Y.; Guerrero, I.; Hidalgo, A.; Taylor, A.; Whittle, J.R.S.; Ingham, P.W.
Nature 341, 508-513, 1989
A:Title: A protein with several possible membrane-spanning domains encoded by the Drosoph
A:Reference number: S06119; MUID:90015164; PMID:2797178
A:Accession: S06119
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-1299 <NAK>
A:Cross-references: UNIPROT:P18502; UNIPARC:UPI000014EA18; GB:X17558; NID:g8389; PIDN:CA
R:Hooper, J.E.; Scott, M.P.
Cell 59, 751-765, 1989
A:Title: The Drosophila patched gene encodes a putative membrane protein required for se
A:Reference number: A33468; MUID:90058658; PMID:2582494
A:Accession: A33468
A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
A:Molecule type: mRNA
A:Residues: 1-110, 'R', 112-273, 'G', 275-331, 'R', 333-635, 'P', 637-861, 'DVF', 878, 'Y', 880-1299
A:Cross-references: UNIPARC:UPI000016BCA3; GB:M28418; GB:M28999; NID:g552097; PID:g55209
C:Genetic: A:Gene: FlyBase:ptc
A:Cross-references: FlyBase:FBgn0003892
A:Map position: 2 4AD3-D4
C:Superfamily: Drosophila membrane protein patched
C:Keywords: glycoprotein; transmembrane protein
F:74-92/Domain: transmembrane #status predicted <TM01>
F:427-448/Domain: transmembrane #status predicted <TM02>
F:456-503/Domain: transmembrane #status predicted <TM03>
F:529-555/Domain: transmembrane #status predicted <TM04>
F:557-585/Domain: transmembrane #status predicted <TM05>
F:677-699/Domain: transmembrane #status predicted <TM06>
F:967-1017/Domain: transmembrane #status predicted <TM07>
F:1019-1047/Domain: transmembrane #status predicted <TM08>
F:1061-1086/Domain: transmembrane #status predicted <TM09>
F:1099-1121/Domain: transmembrane #status predicted <TM10>
F:142,298,335,388,807,861,1199,1271/Binding site: carbohydrate (Asn) (covalent) #status
Query Match 6.7%; Score 464.5; DB 1; Length 1299;
Best Local Similarity 20.4%; Pred. No. 4,4e-24;
Matches 245; Conservative 174; Mismatches 472; Indels 311; Gaps 41;
Qy 321 KETSLSDKTSFSTHTLLGPFQGWGMVMSWPTLLIVSVIRPVNLAALGVTELTDDV 380
Db 43 KKKARSSRAIYLRYSVQSHLETLTGSSVQKAGKVLFAIIVLSTFCVCLKSAQHSKVH 102
Qy 381 EIMSAFNSQARSEKAFNDHFG-PFRITNOVILTA---PNRSSYRDSLLGPKNPSGIL 436
Db 103 QLMIGGGGLLEALATYQKTIGEDBSATHQLLIQTHDPAASVLAHQALL-----A 153
Qy 437 DLDLLELLEDRNLHLYVSPEDARNTSLDICTAALNPDSITLY-----DCC 486
Db 154 HLEVLVAKATVAKVNLVDTF-W-----GLRDMCMPSTSPSEGIYUIEQLIRHLIPCS 204
Qy 487 INSLLOYFN-----NRTLLLLTAN-----OTLMGQSQVUMKQH 521
Db 205 ITPPLDCFWEGSQLLGPESAAVVIPIGLNQRLLMTLLNPASVQMYMKQMSBEKISFDFTV 264

Qy 522 FLYCANAPLTERKDGALALSCM-----ADYCAPVPPLAIG--GYK 560
Db 265 EGYMRAAIA---SGYMKRPPCLINPLNCPPTAPNKNSTOPPDVGA---ILSGCYGYA 317
Qy 561 GKDYEAEALMTFTSLNYPAGDPLTAQ----- 589
Db 318 AAKMIMPEBLIVGAKRR-RSGHLRKAQALQSVOVQMTKEKMYDQMNYKVHNGTQE 376
Qy 590 -----KLMEBAF---LEEMRAFORRMAGMFOVTFPABSLDEINRTTAEDLPFATSY 640
Db 377 KAAEVLNMQNRNFSREVEQLLRKQSRINQNDIYVSSAALDDLAKSHPSALSVIGV 436
Qy 641 IYIFLYISLALGSYSMSRWVWDSKATGLGGAVALCAVMAKGFSLIGRSLVLQ 700
Db 437 AVTVLY---AFCTLLRW-RDPVRCQSSVAGVGLIMCFSTYAGLGLSALLGIVFNASTQ 492
Qy 701 VVPELVLSGADNIFIFLEVQRLPRRGRGEFREVHIGALRVAPSMGLSSEALICFL 760
Db 493 VVPELALGLVDHIFMLTAAVAESNRR--EQTKL---ILKKVGSILFSACTAGSFPA 546
Qy 761 GALTMPAVRTPALTSGLAVILDFLOMSAFVALSLDSKQEASRLDV-CCCV-----K 814
Db 547 AAFIVPALIKVFCLOALVMSGNLAALLVPRAMISLDRRTGRADIIFCCCFVWKEQ 606
Qy 815 PQELRP-----PQSGEL-----LIGPF 832
Db 607 PKVAPPVLPLNNNGRGARHPKSCNNNRVALPAQNPLEGRADIPGSSHSIASFSLATFA 666
Qy 833 OKAYAPPELLMHTTRVNLPLALPGVSLYEMCHISVGLDEMLPKSYLDYFLPNR 892
Db 667 FQHYVPELRMSVVKLTVMGFLALISLVASTRQDGLDIIIVPKOSHEHKEFLDAQTR 726
Qy 893 YEVGAPVYVFTLGYNFSSEKGMNAICSSAGCNFSFTOKIQTATEPEQSILAIP--- 949
Db 727 LF--GFYGMVANTQ-----NFEVPTQQLLRQV-HDSFRAVPHVI 764
Qy 950 -----ASSVVDIFDWLTPBSSCCRLYISGBNKKXFCPSIVNSLNCILKNC----- 993
Db 765 KNDNGCLPDPFWLLLFSEWL-----GNLQKIFDEERYDRGLTKECWFPNASSDA 812
Qy 994 -----MSITMGSVRSVQ-----FHKYL--PWFLDGRPNIKCPKGG 1028
Db 813 ILAYVLIVQTHGVNDPVKVELVLTNRVNSDGIINORAFYNLSAM---ATMSSPTEL 868
Qy 1029 LAA--YSTSVNLTSQGVLAERFMAVHPLKNSOD-----YTEALRAAR 1070
Db 869 LRANCIARRRANGASGCKLYPERQYFHQ--NEYDLKIPKSLPVYAQMPLYHGLDPTS 926
Qy 1071 ELAANITADLRKVPCTDPAFEV--FPTYITNVFYEQYLITLPEGLFMLSCLVPTFAVSC 1128
Db 927 QI-KTLIGHIRDLVSKYEGFGILPNYPSGIPFIWFQWYT-LRSSIAMILACVLIALVLV 984
Qy 1123 LLLGLDASGLNLSTYMWLVDTYGFMAVLDISNAVSLNVNVAQMSVEFVSHTRS 1188
Db 985 SLLLSLWAAVLYLVISLASLAQIFGANTLLGIKISALPAVILLISVMMKCF--NVLIS 1042
Qy 1189 FAISTKPTWLERAKATISMGSAAVPAVAMTNLPGLIVLGLAKAQLIQIPEFRNLITLL 1248
Db 1043 LGPMTSVGNRQKRVLDLSQMSLGRPLVHGMLTSGAVFPLTSSPFEVLRHCKMLLVLC 1102
Qy 1249 LGLHGLVFLPYLISYVGPVNPALALFQKRAEEVAAMVASCPNHSRVSADNIYVN 1308
Db 1103 VGACNSLIVFPIILSMVPEAE-LVPLF-----HDIRISTSPPLVR 1143
Qy 1309 HS 1310
Db 1144 SS 1145

Search completed: March 22, 2006, 22:56:20
Job time : 63 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: March 22, 2006, 22:47:27 ; Search time 265 Seconds
(without alignments)
3546.278 Million cell updates/sec

Title: US-10-736-769-4
Perfect score: 6909
Sequence: 1 M6A6RGKMLMALLRLAQ.....GSIKAGATSNFLPNNRQP 1332

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_05.80: *
1: uniprot_sprot: *
2: uniprot_trembl: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6872.5	99.5	1359	1 NPCL1_HUMAN	Q9UHC9 homo sapien
2	5421.5	78.5	1331	1 NPCL1_RAT	Q6C3U3 rattus norv
3	5407	76.3	1333	1 NPCL1_MOUSE	Q6C3U4 mus musculu
4	3179	46.0	1132	2 Q4T749_TITING	Q4T749 tetradon n
5	2413	34.9	1277	2 Q9JIG3_CRIGR	Q9JIG3 cricetulus
6	2412	34.9	1277	1 NPCL1_PIG	P55941 sus scrofa
7	2405	34.8	1276	2 Q8WKD8_FELCA	Q8WKD8 felis silve
8	2404.5	34.8	1276	2 Q9GK52_CANFA	Q9GK52 canis famli
9	2402.5	34.8	1276	1 NPCL1_HUMAN	O15118 homo sapien
10	2402	34.8	1289	2 Q59GR1_HUMAN	Q59GR1 homo sapien
11	2400	34.7	1277	2 Q7TMD4_MOUSE	Q7TMD4 mus musculu
12	2399	34.7	1276	2 Q9N000_FELCA	Q9N0Q0 felis silve
13	2395	34.7	1276	2 Q8W149_FELCA	Q8W149 felis silve
14	2385	34.5	1278	1 NPCL1_MOUSE	O35604 mus musculu
15	2367.5	34.3	1277	2 Q9GLC9_BOVIN	Q9GLC9 bos taurus
16	2361	34.2	1286	2 Q9T775_RABIT	Q9T775 oryctolagus
17	2338	33.7	1209	2 Q4RWY5_TITING	Q4RWY5 tetradon n
18	2262	32.7	1287	2 Q9U5M1_DROME	Q9U5M1 drosophila
19	2256	32.7	1287	2 Q9U5M1_DROME	Q9U5M1 drosophila
20	2253	31.6	1287	2 Q7YU59_DROME	Q7YU59 drosophila
21	2205	31.9	1291	2 Q7Q409_ANOGA	Q7Q409 anopheles g
22	1864.5	27.0	1233	2 Q9VRC9_DROME	Q9VRC9 drosophila
23	1760	25.5	1003	2 Q7PS03_ANOGA	Q7PS03 anopheles g
24	1725	25.0	1075	2 Q5INK7_MAGGR	Q5INK7 magnaporthe
25	1712.5	24.8	1361	2 Q7XUB7_ORISA	Q7XUB7 oryza sativ
26	1700.5	24.6	1271	2 Q5B8G1_EMBENI	Q5B8G1 aspergillus
27	1659.5	24.0	1330	2 Q5KGS9_CRYNE	Q5KGS9 cryptococcu
28	1658.5	24.0	1330	2 Q5SSD4_CRYNE	Q5SSD4 cryptococcu
29	1643	23.8	1273	2 Q4MMG5_ASPFU	Q4MMG5 aspergillus
30	1614.5	23.4	1295	2 Q410K4_GIBZE	Q410K4 gibberella
31	1549	22.4	1264	2 Q6BT03_DEBHA	Q6BT03 debaryomyce

ALIGNMENTS

RESULT 1	NPCL1_HUMAN	STANDARD	PRT: 1359 AA.	
ID	NPCL1_HUMAN	Q9UHC9; Q6R3Q4; Q9UHC8;		
AC	Q9UHC9; Q6R3Q4; Q9UHC8;			
DT	13-SEP-2005 (Rel. 48, Created)			
DT	13-SEP-2005 (Rel. 48, Last sequence update)			
DT	13-SEP-2005 (Rel. 48, Last annotation update)			
DE	Niemann-Pick C1-like protein 1 precursor.			
GN	Name=NPCL1;			
OS	Homo sapiens (human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae;			
OC	Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	NCULEOTIDE SEQUENCE [MRNA] (ISOFORMS 1; 2 AND 3), AND TISSUE SPECIFICITY.			
RP	MEDLINE=20247253; PubMed=10783261; DOI=10.1006/geno.2000.6151;			
RA	Davies J.P., Levy B., Ioannou Y.A.,			
RT	"Evidence for a Niemann-Pick C (NPC) gene family: identification and characterization of NPCL1."			
RL	Genomics-65:137-145(2000).			
RL	[2]			
RN	NCULEOTIDE SEQUENCE [MRNA] (ISOFORM 2), AND TISSUE SPECIFICITY.			
RP	PubMed=15976318; DOI=10.1126/science.1093131;			
RA	Altman S.W., Davis H.R., Jr., Zhu L.-U., Yao X., Hoos L.M.,			
RT	Terzileff G., Iyer S.P.N., Maguire M., Golovko A., Zeng M., Wang L.,			
RL	"Niemann-Pick C1 like 1 protein is critical for intestinal cholesterol absorption."			
RL	Science 303:1201-1204(2004).			
RL	[3]			
RP	TISSUE SPECIFICITY, AND SUBCELLULAR LOCATION.			
RP	PubMed=15671032; DOI=10.1074/jbc.M409110200;			
RA	Davies J.P., Scott C., Oishi K., Liepis A., Ioannou Y.A.;			
RT	"Inactivation of NPCL1 causes multiple lipid transport defects and protects against diet-induced hypercholesterolemia."			
RL	J. Biol. Chem. 280:12710-12720(2005).			
RL	[4]			
RP	INDUCTION.			
RP	PubMed=15504518; DOI=10.1194/jlr.M400400-JLR200;			
RA	van der Veen J.N., Krull J.K., Havinga R., Baller J.F.W., Chimini G.,			
RT	Lestavel S., Staelen B., Groot P.H.E., Groen A.K., Kuipers F.;			
RT	"Reduced cholesterol absorption upon PPARdelta activation coincides with decreased intestinal expression of NPCL1."			
RL	J. Lipid Res. 46:526-534(2005).			
RL	[5]			
RP	FUNCTION.			
RP	PubMed=15528087; DOI=10.1073/pnas.0500269102;			
RA	Garcia-Calvo M., Lienock J., Bull H.G., Haves B.E., Burnett D.A.,			
RA	Braun M.P., Crona J.H., Davis H.R., Jr., Dean D.C., Detmers P.A.,			
RA	Gratianno M.P., Hughes M., Macintyre D.B., Ogawa A., O'Neill K.A.,			
RA	Iyer S.P.N., Shevell D.E., Smith M.M., Tang Y.S., Makarewicz A.M.,			
RA	Ujjainwalla F., Altman S.W., Chapman K.T., Thornberry N.A.;			

32	1527.5	22.1	1275	2	Q9SHN9_ARATH	Q9SHN9 arabidopsis
33	1511.5	21.9	1342	2	Q9TVK6_DICDI	Q9TVK6 dictyosteli
34	1508	21.8	1239	2	Q6CBAL_YARLI	Q6CBAL yarrowia li
35	1500	21.7	1162	2	Q7RML9_NEUCR	Q7RML9 neurospora
36	1461.5	21.2	1256	2	Q59ZV0_CANAL	Q59ZV0 candida alb
37	1452.5	21.0	1397	2	Q551C5_DICDI	Q551C5 dictyosteli
38	1437.5	20.8	1489	2	Q4PE83_USRMA	Q4PE83 ussilaago ma
39	1394	20.2	1055	2	Q9SVF0_ARATH	Q9SVF0 arabidopsis
40	1386.5	20.1	620	2	Q4H344_CIOIN	Q4H344 clona intes
41	1359	19.7	1382	2	Q618V4_CAEBR	Q618V4 caenorhabd
42	1341.5	19.4	1383	1	NPCL1_CAEEL	NPCL127 caenorhabd
43	1329	19.2	1170	2	Q122D0_YEAST	Q122D0 saccharomyc
44	1323	19.1	1339	2	Q5ORB6_ENTHI	Q5ORB6 entamoeba h
45	1312.5	19.0	1178	2	Q750G1_ASHGO	Q750G1 ashbya goes

RT "The target of ezetimibe is Niemann-Pick C1-like 1 (NPC1L1).";
 RL Proc. Natl. Acad. Sci. U.S.A. 102:8132-8137 (2005).
 RN [6]
 RP VARIANTS LEU-55 AND ASN-1233.
 RX PubMed=15679830; DOI=10.1111/j.1399-0004.2004.00388.x;
 RA Wang J., Williams C.M., Hegele R.A.; 1399-0004.2004.00388.x;
 RT "Compound heterozygosity for two non-synonymous polymorphisms in
 NPC1L1 in a non-responder to ezetimibe.";
 RL Clin. Genet. 67:175-177 (2005).
 CC -1- FUNCTION: Play a major role in cholesterol homeostasis. Is
 critical for the uptake of cholesterol across the plasma membrane
 of the intestinal enterocyte. Is the direct molecular target of
 ezetimibe, a drug that inhibits cholesterol absorption. Lack of
 activity leads to multiple lipid transport defects. The protein
 may have a function in the transport of multiple lipids and their
 homeostasis, and may play a critical role in regulating lipid
 metabolism.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Subfractionation
 of brush border membranes from proximal enterocytes suggests
 considerable association with the apical membrane fraction. Exists
 as a predominantly cell surface membrane expressed protein (By
 similarity). According to Ref.3 localizes in a subcellular
 vesicular compartment rich in RAB5.
 CC -1- ALTERNATIVE PRODUCTS:
 Event=Alternative splicing; Named isoforms=3;
 Name=1;
 IsoId=Q9UHC9-1; Sequence=Displayed;
 Name=2; Synonyms=NPC1L1DELTAEL5;
 IsoId=Q9UHC9-2; Sequence=VSP_015314;
 Name=3; Synonyms=NPC1L1;
 IsoId=Q9UHC9-3; Sequence=VSP_015312; VSP_015313;
 CC -1- TISSUE SPECIFICITY: Widely expressed. Expressed in liver. Also
 expressed in small intestine, pancreas, kidney, lung, pancreas,
 spleen, heart, gall bladder, brain, testis, stomach and muscle.
 CC -1- INDUCTION: Expression is decreased in Caco-2 cells upon PPAR
 activation.
 CC -1- PTM: Highly glycosylated (By similarity).
 CC -1- POLYMORPHISM: Variations in NPC1L1 gene could be associated with
 nonresponse to ezetimibe treatment.
 CC -1- SIMILARITY: Belongs to the patched family.
 CC -1- SIMILARITY: Contains 1 SSD (sterol-sensing) domain.
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 use as long as its content is in no way modified and this statement is not
 removed.
 CC
 CC EMBL; AF192522; AAF20396.1; -; mRNA.
 DR EMBL; AF192523; AAF20397.1; -; mRNA.
 DR EMBL; AY15256; AAS56939.1; -; mRNA.
 DR EMBL; AY437865; AAR97886.1; -; mRNA.
 DR Ensembl; ENSG00000015520; Homo sapiens.
 DR HGNC; HGNC:7898; NPC1L1.
 DR MIM; 608010; -.
 DR GO; GO:0015918; P:sterol transport; TAS.
 DR InterPro; IPR004765; NP_C type.
 DR InterPro; IPR003392; Patched.
 DR InterPro; IPR000731; SSD 5TM.
 DR Pfam; PF02460; Patched; 1.
 DR TIGRFAMs; TIGR00917; ZAO60601; 1.
 DR PROSITE; PS50156; SSD; 1.
 DR Alternative splicing; Cholesterol metabolism; Glycoprotein;
 lipid metabolism; Polymorphism; Signal; Steroid metabolism;
 Transmembrane.
 KW SIGNAL
 FT 1 21 Potential.
 FT CHAIN 22 1359 Niemann-Pick C1-like protein 1.
 FT TOPO_DOM 22 284 Extracellular (Potential).
 FT TRANSMEM 285 305 1 (Potential).
 FT TOPO_DOM 306 351 Cytoplasmic (Potential).
 FT TRANSMEM 352 372 2 (Potential).
 FT TOPO_DOM 373 632 Extracellular (Potential).
 FT TRANSMEM 633 653 3 (Potential).

FT	TOPO_DOM	654	666	Cytoplasmic (Potential).
FT	TRANSMEM	667	697	4 (Potential).
FT	TOPO_DOM	668	696	Extracellular (Potential).
FT	TRANSMEM	697	717	5 (Potential).
FT	TOPO_DOM	718	742	Cytoplasmic (Potential).
FT	TRANSMEM	743	763	6 (Potential).
FT	TOPO_DOM	764	776	Extracellular (Potential).
FT	TRANSMEM	777	797	7 (Potential).
FT	TOPO_DOM	798	846	Cytoplasmic (Potential).
FT	TRANSMEM	847	867	8 (Potential).
FT	TOPO_DOM	868	882	Extracellular (Potential).
FT	TRANSMEM	883	903	9 (Potential).
FT	TOPO_DOM	904	1139	Cytoplasmic (Potential).
FT	TRANSMEM	1140	1160	10 (Potential).
FT	TOPO_DOM	1161	1189	Extracellular (Potential).
FT	TRANSMEM	1169	1189	11 (Potential).
FT	TOPO_DOM	1190	1191	Cytoplasmic (Potential).
FT	TRANSMEM	1192	1212	12 (Potential).
FT	TOPO_DOM	1213	1236	Extracellular (Potential).
FT	TRANSMEM	1237	1257	13 (Potential).
FT	TOPO_DOM	1258	1359	Cytoplasmic (Potential).
FT	DOMAIN	632	797	SSD.
FT	COMPBIAS	500	503	Poly-Leu.
FT	CARBOHYD	54	54	N-linked (GlcNAc. . .) (Potential).
FT	CARBOHYD	132	132	N-linked (GlcNAc. . .) (Potential).
FT	CARBOHYD	138	138	N-linked (GlcNAc. . .) (Potential).
FT	CARBOHYD	244	244	N-linked (GlcNAc. . .) (Potential).
FT	CARBOHYD	416	416	N-linked (GlcNAc. . .) (Potential).
FT	CARBOHYD	431	431	N-linked (GlcNAc. . .) (Potential).
FT	CARBOHYD	464	464	N-linked (GlcNAc. . .) (Potential).
FT	CARBOHYD	479	479	N-linked (GlcNAc. . .) (Potential).
FT	CARBOHYD	497	497	N-linked (GlcNAc. . .) (Potential).
FT	CARBOHYD	506	506	N-linked (GlcNAc. . .) (Potential).
FT	CARBOHYD	626	626	N-linked (GlcNAc. . .) (Potential).
FT	VARSPLIC	723	724	RL -> GP (in isoform 3).
FT	VARSPLIC	725	1359	/FTId=VSP_015312.
FT	VARSPLIC	1046	1072	Missing (in isoform 3).
FT	VARSPLIC	1233	1233	/FTId=VSP_015313.
FT	VARIANT	55	55	Missing (in isoform 2).
FT	VARIANT	1233	1233	/FTId=VSP_015314.
FT	VARIANT	1233	1233	V -> L (nonresponse to ezetimibe treatment).
FT	VARIANT	1233	1233	/FTId=VAR_023369.
FT	VARIANT	1233	1233	I -> N (nonresponse to ezetimibe treatment).
FT	VARIANT	1233	1233	/FTId=VAR_023370.
SO	SEQUENCE	1359 AA; 148698 MW; 32250530938768B6 CRC64;		

Query Match 99.5%; Score 6872.5; DB 1; Length 1359;
 Best Local Similarity 97.9%; Pred. No. 0;
 Matches 1330; Conservative 0; Mismatches 2; Indels 27; Gaps 1;

QY	1	MAEAGRGRLWALLRLAQAQEPYTTIHOPGCAFYDECGKNPELSGSLMTSNVSCSN	60
DB	1	MAEAGRGRLWALLRLAQAQEPYTTIHOPGCAFYDECGKNPELSGSLMTSNVSCSN	60
QY	61	TPARKITGDHLILLOKICPRLYTGPNTOACCSAKQVLSBASLITKALLTRCPACSNF	120
DB	61	TPARKITGDHLILLOKICPRLYTGPNTOACCSAKQVLSBASLITKALLTRCPACSNF	120
QY	121	VNLHGNTCSPNQSLFIVNTRVAQAGALPAVVAYEAFOHSPAEQSYDSCSRVRVPA	180
DB	121	VNLHGNTCSPNQSLFIVNTRVAQAGALPAVVAYEAFOHSPAEQSYDSCSRVRVPA	180
QY	181	ATLAVGTMGVYGSALCNARWLNFQSGTNGGLAPLDTFPHLBEQAVGSGIQPLNCGV	240
DB	181	ATLAVGTMGVYGSALCNARWLNFQSGTNGGLAPLDTFPHLBEQAVGSGIQPLNCGV	240
QY	241	ARCNSQGDVATCSQCDCAASCPAIPAPALDSTFYIGOMGSLVLIILICSVPAVVTI	300
DB	241	ARCNSQGDVATCSQCDCAASCPAIPAPALDSTFYIGOMGSLVLIILICSVPAVVTI	300
QY	301	ILVGRVAPARDKSRMDPKGTSLSDKLSFSTHTLLQDFQMGCTWASNPDLTLVLVS	360

[illegible]

NCPI1_RAT
ID NCPI1_RAT STANDARD; PRT; 1331 AA.
AC Q6TJ03;
DT 13-SEP-2005 (Rel. 48, Created)
DT 13-SEP-2005 (Rel. 48, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Niemann-Pick C1-like protein 1 precursor.
GN Name=NCPI1;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Rattus.
OX NCPI_TaxID=10116;
RN
RP NUCLEOTIDE SEQUENCE [MRNA] , TISSUE SPECIFICITY, AND SUBCELLULAR
RP LOCATION.
RC STRAIN=Sprague-Dawley;
RX PubMed=14976318; DOI=10.1126/science.1093131;
RA Altman S.W., Davis H.R. Jr., Zhu L.-J., Yao X., Hoos L.M.,
RA Tetzloff G., Iyer S.P.N., Maguire M., Golovko A., Zeng M., Wang L.,
RA Murgolo N., Graziano M.P.;
RT "Niemann-Pick C1 like 1 protein is critical for intestinal cholesterol
RT absorption.";
RL Science 303:1201-1204(2004).
RN
RN [2]
RN SUBCELLULAR LOCATION, AND CHARACTERIZATION.
RP PubMed=15777641; DOI=10.1016/j.bbaen.2004.12.021;
RX Iyer S.P.N., Yao X., Crona J.H., Hoos L.M., Tetzloff G.,
RA Davis H.R. Jr., Graziano M.P., Altman S.W.;
RT "Characterization of the putative native and recombinant rat sterol
RT transporter Niemann-Pick C1 like 1 (NPC1L1) protein.";
RL Biochim. Biophys. Acta 1722:282-292(2005).
RN
RN [3]
RN FUNCTION.
RP PubMed=15928087; DOI=10.1073/pnas.0500269102;
RX Garcia-Calvo M., Llenock J., Bull H.G., Hawes B.E., Burnett D.A.,
RA Braun M.P., Crona J.H., Davis H.R. Jr., Dean D.C., Detmers P.A.,
RA Graziano M.P., Hughes M., MacIntyre D.E., Ogawa A., O'Neill K.A.,
RA Iyer S.P.N., Shevell D.E., Smith M.M., Tang Y.S., Makarewicz A.M.,
RA Ulfwallia F., Altman S.W., Chapman K.T., Thorberry N.A.;
RT "The target of ezetimibe is Niemann-Pick C1-like 1 (NPC1L1).";
RL Proc. Natl. Acad. Sci. U.S.A. 102:8132-8137(2005).
RN
RN [1] FUNCTION: Play a major role in cholesterol homeostasis. Is
CC critical for the uptake of cholesterol across the plasma membrane
CC of the intestinal enterocyte. Is the direct molecular target of
CC ezetimibe, a drug that inhibits cholesterol absorption (By
CC similarity). Lack of activity leads to multiple lipid transport
CC defects. The protein may have a function in the transport of
CC multiple lipids and their homeostasis.
CC
CC [1] SUBCELLULAR LOCATION: Integral membrane protein. Subfractionation
CC of brush border membranes from proximal enterocytes suggests
CC considerable association with the apical membrane fraction. Exists
CC as a predominantly cell surface membrane expressed protein.
CC
CC [1] TISSUE SPECIFICITY: Small intestine showed the highest level of
CC expression. Expression in other tissue including gall bladder,
CC liver, testis, and stomach is also observed. Along the duodenum-
CC ileum axis, the levels vary in different segments of intestine,
CC with peak expression in the proximal jejunum. Protein expression
CC is confined to the enterocyte. Discrete localization to the
CC epithelial layer bordering the luminal space along the crypt-
CC villus axis. Protein expression in the enterocyte is observed
CC closest to the luminal space. Expression in enterocytes from the
CC proximal (jejunum) but not in the distal (ileum) region.
CC
CC [1] PTM: Highly glycosylated.
CC
CC [1] SIMILARITY: Belongs to the patched family.
CC [1] SIMILARITY: Contains 1 SSD (sterol-sensing) domain.
CC
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.

DR	EMBL: AY437867; AAR97888.1; -; mRNA.
DR	RGD: 1303135; Npc1l1.
DR	InterPro: IPR004765; NP_C_type.
DR	InterPro: IPR003392; Patched.
DR	InterPro: IPR000731; SSD 5TM.
DR	Pfam: PF02460; Patched; 1.
DR	TIGRFAM: TIGR00917; 2A060601; 1.
DR	PROSITE: PS50156; SSD: 1.
KW	cholesterol metabolism; Glycoprotein; Lipid metabolism; signal; steroid metabolism; Transmembrane.
FT	SIGNAL 1 20
FT	CHAIN 1 1331
FT	TOPO_DOM 21 282
FT	TRANSMEM 283 303
FT	TOPO_DOM 304 352
FT	TRANSMEM 353 373
FT	TOPO_DOM 374 632
FT	TRANSMEM 633 653
FT	TOPO_DOM 654 665
FT	TRANSMEM 666 686
FT	TOPO_DOM 687 696
FT	TRANSMEM 697 717
FT	TOPO_DOM 718 742
FT	TRANSMEM 743 763
FT	TOPO_DOM 764 776
FT	TRANSMEM 777 797
FT	TOPO_DOM 798 846
FT	TRANSMEM 847 867
FT	TOPO_DOM 868 1113
FT	TRANSMEM 1114 1134
FT	TOPO_DOM 1135 1142
FT	TRANSMEM 1143 1163
FT	TOPO_DOM 1164 1165
FT	TRANSMEM 1166 1187
FT	TOPO_DOM 1187 1206
FT	TRANSMEM 1207 1237
FT	TOPO_DOM 1238 1242
FT	TRANSMEM 1243 1263
FT	TOPO_DOM 1264 1331
FT	DOMAIN 632 797
FT	CARBOHYD 53 53
FT	CARBOHYD 85 85
FT	CARBOHYD 138 138
FT	CARBOHYD 244 244
FT	CARBOHYD 416 416
FT	CARBOHYD 431 431
FT	CARBOHYD 464 464
FT	CARBOHYD 479 479
FT	CARBOHYD 497 497
FT	CARBOHYD 506 506
FT	CARBOHYD 606 606
FT	CARBOHYD 626 626
FT	CARBOHYD 909 909
FT	CARBOHYD 917 917
FT	CARBOHYD 996 996
FT	CARBOHYD 1038 1038
FT	CARBOHYD 1076 1076
FT	CARBOHYD 1331 AA; 146415 MW; 2E10EF2E3A337F70 CRC64;
SEQ	SEQUENCE
Query Match	78.5%; Score 5421.5; DB 1; Length 1331;
Best Local Similarity	77.8%; Pred. No. 0;
Matches 1039;	Conservative 120; Mismatches 169; Indels 7; Gaps 6;
QY	1 MAEAGIRGWLIALWLLRLAQLSEPYTTIHQGYCAFYDECGKNPELGSIMTLNVNCSLSN
DB	1 MAAMWL-GWLWALLLSAAGELVTPKHEAGVCTFEBCGKNPELSGGLTSLNSVCSLSN
QY	61 TPARKITGDHLTLLOKICPRLTYGENTQ-ACCSAKOLVLEASLSITKALLTRCPACSDN
DB	60 TPARKVHTGHLTLLOKICPRLTYGENTQ-ACCSAKOLVLEASLSITKALLTRCPACSDN
QY	120 FVNLCHNTGSPNOSLFINTVTRVAOLGAGOLPAVVAVYEAFFQHSFPEQSYDSCSRVPA
DB	120 FVNLCHNTGSPNOSLFINTVTRVAOLGAGOLPAVVAVYEAFFQHSFPEQSYDSCSRVPA

DB	120 FVSLHCHNTGSPDOSLFINTVTRVVERGAGEPPAVVAVYAFQORSPAEKAYESCQVRIPA
QY	180 AATLAVGTMGVYGSALCNAQRMWLNFGQDGTNGALAPLDTTHLLPEGAVSGIQLNLEG
DB	180 AASLAVGSMGVYGSALCNAQRMWLNFGQDGTNGALAPLDTTHLLPEGALPDGIQPLNGK
QY	240 VARNESGQDDVATGSCDDCAASCALARPQALDSTFLGCMPSGLVLIILLCSPAVVT
DB	240 IAPCNESQDDSAVSCDDCAASCALARPQALDSTFLGCMPSGLVLIILLCSPAVVT
QY	300 ILVGFRAVPARADSKQWDPKKGTSLSDKLSFSTHTLLGQFFQGGTVMASWPLTVLS
DB	300 AVLVALRVSRNKNKAKGAPQEARPLPHKHLSPHTLIGRFQWGTNVASWPLTVLNL
QY	360 VLPVVALAAGVETLETTPDVELWSAPNSQARSEKAFDHPGPPFRNTQVLTAPNRS
DB	360 FIVVIALAAGVETLETTPDVELWSAPNSQARSEKAFDHPGPPFRNTQVLTAPNRS
QY	420 YRYDILLGPKNFGSLDLDDLLELLEQERLRLQVWSPEAKRNI SLQDICAPLNPDN
DB	420 KYDILLGSKNFGSLDLDDLLELLEQERLRLQVWSPEAKRNI SLQDICAPLNPYN
QY	480 TSLVCCINSLLQYQNNRTLLLTANOTLNGQTSQVDMKHFLYCANAPLTFKDGTLA
DB	480 TSLVCCNSLLQYQNNRTLLLTANOTLNGQTSQVDMKHFLYCANAPLTFKDGTLA
QY	540 LSCMDYGAPEVFPFLAIGYKGDYSEAEALIMFESLNNYPAGDBRLAQAALMEAPLE
DB	540 LSCMDYGAPEVFPFLAIGYKGDYSEAEALIMFESLNNYPADDBRMAQAALMEAPLE
QY	600 MRAPGRMAQPOVFTFMRSLDEINTTMDLPFTFSTYIVFLYLSLALGYSMSR
DB	600 MESFRANSDRFQVAFSERSLEDEINTTQDLPVFVASYIVFLYLSLALGYSRCSR
QY	660 VVVDKATLIGGVAVVVGAVMAAMGFPSYIGIRSSLYILQVPELVLSVGDNFIPIVL
DB	660 VAVESKATLIGGVAVVVGAVLAAAGFYSYIGVBSLYIIVPELVLAAGDNFIPIVL
QY	720 EYQRLPRRPGREVHIGRALGRVAPSMLLCSLSBAICFFLGAALTMPAVRTFALTSLA
DB	720 EYQRLPRRPGREAHIGRTLIGSVAPSMLLCSLSBAICFFLGAALTMPAVRTFALTSLA
QY	780 VILDLTLQMSAFVALLSLDSKQEA SRUDVCCVKKPQGLPPPGQEGILLGFQQAIVPF
DB	780 IILDLTLMTAFVALLSLDSKQEA SRDVCSTRLKPPKKEGILLGFPRKIVAPF
QY	840 LHMTRGVLLFLFALRGVSLYSWCHTSVGLDQSLAPKOSYLLDYFLPLNRYEYVAP
DB	840 LHMTRGVVLLFLFALRGVSLYSWCHTSVGLDQSLAPKOSYLLDYFLPLNRYLEVGP
QY	900 VYFVTTLGYNFSSEAGMNAICSSAGCNNSFTOKIQVATPEPESQYLAIPASSWVDIFD
DB	900 VYFVTTSGFNSSEAGMNAICSSAGCSFSLQKIQVASEPDSQYVLAIASSWVDIFD
QY	960 WLTTP-SGCCRLYISGPNQKFCPTVNSLNCXKCMSTITMSVAPSVQGFHKYLPWFIND
DB	960 WLTTPSSSCCRIRYIRPHDCEPSTDSFNCLKCMNNTLIPVPTAQGFHKYLPWFIND
QY	1019 RPNICPKPGGLAAYSTVNLTSDGOVLASRFMAVYKPLKNSODYTEALRAARELANITTA
DB	1020 PPNICPKPGGLAAYRTSNLSSDGOVLASQFMAVHKPLRNSODTEALRAARELANITTA
QY	1079 DLKRVPGTDPAFEVPPYITTVFYEQVYLTJLPEGLFMLSCLVPTFAVSCLLGLDLRS
DB	1080 DLKRVPGTDPNPEVPPYITTVFYEQVYLTJLPEGLFMLSCLVPTFAVSCYLLGLDMCSG
QY	1139 LNLTLSTWLVDTVYGFPAALDISYNNVSLINIVASVMSVEFVSHITRSPALSTKPMWL
DB	1140 LNLTLSTWLVDTVIGLAAVWGISNVAVSLNLVLAIVMSVEFVSHITRSPAVSTKPMWL
QY	1199 ERAKATISMSGSAVPAVAMTNLPGLIYVGLAKAQLQIIPFRNLNLTLLGLHGLVFL
DB	1200 ERAKATVPMGSAVPAVAMTNPGLIYVGLAQAQLQIIPFRNLNLTLLGLHGLVFL

QY 1259 PVLSTVGPDPVNPALALBQRAEEAAVAVVAVSCPNHPSRVSTADNIYVNHSEGS1-KG 1317
 DB 1260 PVLSTVGPDPVNPALALBQRAEEAAVAVVAVSCPNHPSRVSTADNIYVNHSEGS1-KG 1316
 QY 1318 AGAISNPLPNNKQF 1332
 DB 1317 ANAARSLPKSDQK 1331

RESULT 3
 NPC1L1_MOUSE STANDARD; PRT; 1333 AA.
 ID NPC1L1_MOUSE
 AC 06T3U4; Q5SVX1;
 DT 13-SEP-2005 (Rel. 48, Created)
 DT 13-SEP-2005 (Rel. 48, Last sequence update)
 DT 13-SEP-2005 (Rel. 48, Last annotation update)
 DE Niemann-Pick C1-like protein 1 precursor.
 GN Name=Npc1l1;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP NUCLEOTIDE SEQUENCE [MRNA], FUNCTION, AND TISSUE SPECIFICITY.
 RC STRAIN=C57BL/6;
 RA PubMed:14976318; DOI=10.1126/science.1093131;
 RA Altman S.W., Davis H.R. Jr., Zhu L.-J., Yao X., Hoos L.M.,
 RA Tetziolf G., Iyer S.P.N., Maguire M., Golovko A., Zeng M., Wang L.,
 RA Murgolo N., Graziano M.P.;
 RA "Niemann-Pick C1 like 1 protein is critical for intestinal cholesterol
 RT absorption.";
 RT Science 303:1201-1204(2004).
 RN [2]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
 RC The mouse genome sequencing consortium;
 RL Submitted (FEB-2005) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP FUNCTION, AND INDUCTION.
 RA PubMed:1517162; DOI=10.1074/jbc.M405817200;
 RA Davis H.R. Jr., Zhu L.-J., Hoos L.M., Tetziolf G., Maguire M., Liu J.,
 RA Yao X., Iyer S.P.N., Lam M.-H., Lund E.G., Delmers P.A.,
 RA Graziano M.P., Altman S.W.;
 RA "Niemann-Pick C1 like 1 (NPC1L1) is the intestinal phytoesterol and
 RT cholesterol transporter and a key modulator of whole-body cholesterol
 RT homeostasis.";
 RT J. Biol. Chem. 279:33586-33592(2004).
 RN [4]
 RP TISSUE SPECIFICITY, AND FUNCTION.
 RA PubMed:15671032; DOI=10.1074/jbc.M409110200;
 RA Davies J.P., Scott C., Oishi K., Liapis A., Ioannou Y.A.;
 RA "Inactivation of NPC1L1 causes multiple lipid transport defects and
 RT protects against diet-induced hypercholesterolemia.";
 RT J. Biol. Chem. 280:12710-12720(2005).
 RN [5]
 RP INDUCTION.
 RA PubMed:15604518; DOI=10.1194/jlr.M400400-0LR200;
 RA van der Veen J.N., Kruit J.K., Havinga R., Baller J.F.W., Chimini G.,
 RA Lestavel S., Staels B., Groot P.H.E., Groen A.K., Kuipers F.;
 RA "Reduced cholesterol absorption upon PPARdelta activation coincides
 RT with decreased intestinal expression of NPC1L1.";
 RT J. Lipid Res. 46:526-534(2005).
 CC -1- FUNCTION: Play a major role in cholesterol homeostasis. Is
 CC critical for the uptake of both phytoesterol and cholesterol across
 CC the plasma membrane of the intestinal enterocyte. Is the direct
 CC molecular target of ezetimibe, a drug that inhibits cholesterol
 CC absorption (By similarity). Lack of activity leads to multiple
 CC lipid transport defects and exerts a protective effect against
 CC diet-induced hyperlipidemia. Lack of activity also causes a
 CC deregulation of CAV1 transport and localization, suggesting that
 CC the observed lipid transport defect may be the indirect result of
 CC an inability of cells to properly target and/or regulate CAV1

CC expression. The protein may have a function in the transport of
 CC multiple lipids and their homeostasis, and may play a critical
 CC role in regulating lipid metabolism.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Subfractionation
 CC of brush border membranes from proximal enterocytes suggests
 CC considerable association with the apical membrane fraction. Exists
 CC as a predominantly cell surface membrane expressed protein.
 CC -1- TISSUE SPECIFICITY: Expressed in small intestine, stomach and
 CC muscle, along with detectable expression in lung, heart, gall
 CC bladder, brain, testis, skin and liver. Expression in liver is
 CC extremely low.
 CC -1- INDUCTION: Cholesterol/choleate feeding resulted in down-regulation
 CC of intestinal expression. Expression is decreased by 35% in the
 CC jejunum upon PPARd activation.
 CC -1- PTM: Highly glycosylated (By similarity).
 CC -1- SIMILARITY: Belongs to the patched family.
 CC -1- SIMILARITY: Contains 1 SSD (sterol-sensing) domain.
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use as long as its content is in no way modified and this statement is not
 CC removed.
 CC -----
 CC EMBL; AY437866; AAR97887.1; -; mRNA.
 CC EMBL; AL607152; CAI24395.1; -; Genomic DNA.
 CC MGI; MGI:2685089; Npc1l1.
 CC GO; GO:0042632; P:cholesterol homeostasis; IMP.
 CC InterPro; IPR004765; NP_C type.
 CC InterPro; IPR003352; Patched.
 CC InterPro; IPR007931; SSD 5TM.
 CC Pfam; PF02460; Patched; 1.
 CC TIGRPFAM; TIGR00917; 2A060601; 1.
 CC PROSITE; PS50156; SSD; 1.
 CC DR Cholesterol metabolism; Glycoprotein; Lipid metabolism; Signal;
 CC Steroid metabolism; Transmembrane.
 CC KM Steroid metabolism; Transmembrane.
 CC FT SIGNAL 1 20
 CC FT CHAIN 21 1333
 CC FT TOPO_DOM 21 284
 CC FT TRANSMEM 285 305
 CC FT TOPO_DOM 306 352
 CC FT TRANSMEM 353 373
 CC FT TOPO_DOM 374 632
 CC FT TRANSMEM 633 653
 CC FT TOPO_DOM 654 665
 CC FT TRANSMEM 666 686
 CC FT TOPO_DOM 687 696
 CC FT TRANSMEM 697 717
 CC FT TOPO_DOM 718 742
 CC FT TRANSMEM 743 763
 CC FT TOPO_DOM 764 776
 CC FT TRANSMEM 777 797
 CC FT TOPO_DOM 798 846
 CC FT TRANSMEM 847 867
 CC FT TOPO_DOM 868 1113
 CC FT TRANSMEM 1114 1134
 CC FT TOPO_DOM 1135 1142
 CC FT TRANSMEM 1143 1163
 CC FT TOPO_DOM 1164 1165
 CC FT TRANSMEM 1166 1186
 CC FT TOPO_DOM 1187 1206
 CC FT TRANSMEM 1207 1227
 CC FT TOPO_DOM 1228 1242
 CC FT TRANSMEM 1243 1263
 CC FT TOPO_DOM 1264 1333
 CC FT DOMAIN 632 797
 CC FT COMPBIAS 267 270
 CC FT COMPBIAS 500 503
 CC FT CONFLICT 476 476
 CC SQ SEQUENCE 1333 AA; 147132 MW; 777152009B552735 CRC64;

Query Match 78.3%; Score 5407; DB 1; Length 1333;
 Best Local Similarity 77.2%; Pred. No. 0;

	Matches	1028:	Conservative	126:	Mismatches	174:	Indels	4:	Gaps	4:
QY	4	AGLRMLLWALLRLA	GEPTTTHQPGY	CAFYDECGKPNBELSGSLMTLSNVSCLSNTPTA	63					
Db	3	AAQGMLLWALLLNSA	QGEGLYTPTHKAF	CFCTFEYECGKNBELSGSLTSLSNISCLSTNPTA	62					
QY	64	RKINGDHIILLOKIC	PELTYTGN-"QA	CCSAKQVLSLEASLSTTALLTRCPACSDNVN	122					
Db	63	RHYVDTHALLQRY	CPRLNGPNDTYACC	STQVLSLSSLSITALLTRCPACSENVVS	122					
QY	123	LHCNHTCSPNOSL	FINTNTRVAOLG	AGOLPAAVAAEAFYQHSFAEOSYOSCSRVYPAAT	182					
Db	123	IHCNHTCSPPDOSL	FINTNTRVYQDRD	GOQLPAAVAAEAFYQHSFAEAYESCSRVYIPAAAS	182					
QY	183	LAVGTCGVYGSAL	CNAORMINFGQD	TNGALPLDITTHLLBPGQAVSGSIOPLNEGVAR	242					
Db	183	LAVSGMCGVYSAL	CNAORMINFGQD	TNGALPLDITTHLLBPGQALADGMKPLDGKITTP	242					
QY	243	CNBSGDDVATCS	CODCAASCALARP	ALDSTFLLGQMPGSLVLIILCSFPAVNTLL	302					
Db	243	CNBSGDEBSAACS	CODCAASCVP	IPPPALRSPFMGRMPLALIIIFTAFFVLLSVL	302					
QY	303	VGFRVAPRADS	KWDPKKGTSLS	OKLSFSTHTLLGQFFQCGFTWASMPLLIYLVIY	362					
Db	303	VYLRVASNRRNKK	KTGSOEAPVLPKRR	KRSPHTVIGRFEPESGTVASMPLLVTLASTIV	362					
QY	363	VVALAAGLVTEL	ITDTPVELMSAPNS	QARSEKAFHDHGPFPFRITNOVILTAPNRSSYRY	422					
Db	363	VIALSVGLTTEL	ITDTPVELMSAPNS	QARKEKAFHDEHGPFPFRITNOVILTAPNRSSYRY	422					
QY	423	DSLLIGPKNFGS	ITDLDLLELLELOER	LRIHQVMSPEAQNRISLQDICTYAPLNDNTSL	482					
Db	423	DSLIGPKNFGS	ITSLDLLELOELLOER	LRIHQVMSHEQRNISIQDICTYAPLNDNTSL	482					
QY	483	YDCCINSLSLOF	QONNRRTLLLTANQ	TNGQTSQVVMKOHFYCANAPLTPKQGTALASC	542					
Db	483	TDCCNSLSLOF	QONNHTLLLTANQ	TNGQTSVLVMKOHFYCANAPLTPYKQGTALASC	542					
QY	543	MADYGA	PVPFPLAIGGYKGDY	SEAEALIMFSLNNYPAGDPRLAQAOLMBEAFLEENRA	602					
Db	543	IADYGAPVFP	FLAVGYGQTDYSEAE	ALIITPSINNPADDPRAHAKLMBEAFLEKMQS	602					
QY	603	FORRAAGNFQ	VTFLERSLDEIN	TTMADLEPIFATSVIVFLYISLAGYSSMSRMVY	662					
Db	603	FORSTADFOJA	FSAEBSERLDEIN	TTTQDLPVFATSVIVFLYISLAGYSSRMSRYAV	662					
QY	663	DSKATLIG	GVAVVVLGA	WMAAGFSPSYGIRSLYILOVPRVLYSVAGDNIFITVLEQ	722					
Db	663	DSKATLIG	DGVAVVVLGA	WMAAGFSPSYGIRSVSLYILOVPRVLYSVAGDNIFITVLEQ	722					
QY	723	RLPRRGP	PREVHIIGRALGR	VAPSMYLSLSLSAICFFIGALPMBDAVTFALTSGIAVIL	782					
Db	723	RLPRRGP	QOREAHIGRTIG	SVAPSMYLSLSLSAICFFIGALTSMAVTFALTSGIAIIF	782					
QY	783	DFLLQMSA	FVALLSLDSK	QEAASRLDVCCVYKPOELPPQGEGLLGFPOKAYAPFLIH	842					
Db	783	DFLLQMTA	FVALLSLDSK	QEAASRPDVCCFSSRNLPKQKEGILLCFEFRKIYTPFLIH	842					
QY	843	WITRGV	ULLFLALFGV	SLYKCHISVGLDQSLAPKXSYLLDYFLPLNRYEYENAPYF	902					
Db	843	RFIRRV	ULLFLVFLFGAN	LYMCNLSVGLDQSLAPKXSYLLDYFLPLNRYLEVPPYF	902					
QY	903	VYTLGYNF	SSSEAKNNAICSS	AGCNFFSFQTKQYATEPEQGYALAI	PASSWVDDIIDLWT	962				
Db	903	DTTSYIN	STEGAKNAICSS	AGCEBSFLTKQIQA	SEFPNGSYVAIAASSWVDDIIDLWT	962				
QY	963	P-SG	CCRLYISGPN	KDPCPSTVNSL	ACLKCMSITMGSVRPSVEQFHXYLPMFLNDRPN	1022				
Db	963	PSSSC	CRITYRGP	PHDEFCBPSTDTSFN	CLCKCMNRBTLAGVRPTTBQFHXYLPMFLNDRPN	1022				
QY	1022	IKCP	PGGLAAVSTVNL	SDQVLA	SRPMAHYKPLKNSODTYEALRAARELAANTADLR	108				
Db	1023	IRCP	PGGLAAVSTVNL	SDGQILASQF	MAHYKPLRNSQDFTFEALRASGLLAANTAEIR	108				

Qy	1082	KVPGCTDAFEFPPPTTINNVPEQYUUTLTPBELPMLSLCVTFPPAVSCULLGLDRLSGILN	1144
Db	1083	KVPPTDPEFVEFPPTTISVNFQOQLTVLPEGIFTLALCFVTPFVVCYLLGLDRLSGILN	11442
Qy	1142	LLSLVIMLIVDPVGVGMALMDISYNVAVSLINLVSAGMSVEFVSHITRSPALSTKPTMLERA	12013
Db	1143	LLSLIMLIVDTIGLMAWGISYNNVSLINLVTANQMSVEFVSHITRSPAVSTKPTRLERA	12014
Qy	1202	KEATISMGSAVAVGAVAMTNLPGLIVLGLAKAQLIQIPFFRLNLLITLGLLHGLVFLPVI	1261
Db	1203	KDATTFMGSAFAFAGVAMTNFPGLILGLFQAQLLIQIPFFRLNLLITLGLLHGLVFLPVI	1262
Qy	1262	LSYVGPDPNPALATEQKRAEBAVAANVWASCPNHPNRSTADNTNYVHNSPEGS-1KGAGA	1332
Db	1263	LSYVGPDPNQALVLEEKLATBA-MWSEBSPCPYPPADANTSDYVNGENPEFIPEINA	13321
Qy	1321	ISNFLPNNGROF 1332	
Db	1332	ASSSLPKSDOKF 1333	
RESULT 4			
Q	Q4T749	TEITNG	PRT; 1132 AA.
ID	Q4T749	TEITNG PRELIMINARY;	
AC	Q4T749		
DT	13-SEP-2005	(TREMBLrel. 31, Created)	
DT	13-SEP-2005	(TREMBLrel. 31, Last sequence update)	
DT	13-SEP-2005	(TREMBLrel. 31, Last annotation update)	
DE	Chromosome undetermined SCAF8317, whole genome shotgun sequence.		
DE	(Fragment).		
GN	ORFname=GSTENG00005930001;		
OS	Tetradodon nigroviridis (Green puffer).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;		
OC	Acanthomorphi; Acanthopterygii; Perccomph; Tetraodontiformes;		
OC	Tetraodontidae; Tetraodontidae; Tetraodon.		
OX	NCHI_TaxID=99883;		
RN	[1]		
RP	NUCLEOTIDE SEQUENCE.		
RA	Jatillon O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,		
RA	Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,		
RA	Micaud S., Jaffe D., Fisher S., Lutellala G., Dossat C., Segurens B.,		
RA	Dasilva C., Salomout M., Levy M., Boudet N., Castellano S.,		
RA	Anthouard V., Jubin C., Castellì V., Karinka M., Vacherie B.,		
RA	Biemont C., Skalli Z., Cattolico L., Poulain J., De Bernardis V.,		
RA	Craud C., Duprat S., Broctier P., Coutanceau J.P., Gouzy J.,		
RA	Kellis G., Lardier G., Chapple C., McKernan K.U., McEwan P., Bosak S.,		
RA	Pavlis M., Wolff J.N., Guigo R., Zody M.C., Mesirov J.,		
RA	Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,		
RA	Lander V., Schachter V., Quetier F., Saurin W., Scarpelli C.,		
RA	Wincker P., Lander E.S., Weissbach J., Roest Crolius H.,		
RT	"Genome duplication in the teleost fish Tetradodon nigroviridis reveals		
RL	the early vertebrate proto-karyotype.";		
RL	Nature 431:946-957(2004).		
RN	[2]		
RP	NUCLEOTIDE SEQUENCE.		
RG	Genoscope, Whitehead Institute Centre for Genome Research;		
RL	Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.		
CC	-1- CAUTION: The sequence shown here is derived from an		
CC	EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is		
CC	preliminary data.		
DR	EMBL; CAAE01008317; CAF91283.1; -; Genomic_DNA.		
FT	NON TER		
SO	SEQUENCE	1132 AA; 125923 MW; 4283CFCA0BBDB97 CRC64;	
Query Match			
Beat Local Similarity 46.0%; Score 3179; DB 2; Length 1132;			
Matches 641; Conservative 167; Mismatches 264; Indels 160; Gaps 11;			
Qy	87	TOACCSAQLVSTASLSITALLTRCAGCNDPNNLHCNHTCSPGNSGLFTNYRVAQL-	145
Db	1	TYACCSINQLSLASLASLISAVILIRCSCADNFAHFICITTCSPQOKTKTKIVTKVNTV	60

Query Match	Score	DB	Length
Similarity 52.0%; Pred. No. 4.3e-216;			
Best Local Match 641; Conservative 167; Mismatches 264; Indels 160; Gaps 11			
87	TOACCAQVLSEASITALLTRCPACDNFVNTHGHNTGCSNQSFLFNTVVAQL	145	
Db 1	TYACCSINQLSLESLALSAKAVILRCSSCADNFHAFICITTCSSQQRKTKIVTAVNT	60	

QY 146 GAGOL--PAVVAVEAFYCHSPAEQSYDSCSRVPAATAATLAVTMCVYGSALCNQRMVL 203
 DB 61 GFDNITREBEVGAVALSTINFPADASFQSCKNVRIPRTGGLAGTMCGRGSHCNQRMV 120
 QY 204 NFOGDTGNGLAPLDTITFHLLEPG--QAVSGIQPLNEGVAKNESGDDVATCSQDCDA 261
 DB 121 DFGDSSNGLAPLDTIDFHLIKPGDTEDELPBGIVPYSGNALKNETTPSGKAKSCQDCDB 180
 QY 262 SCPALARPALDSTFYLQMGPSVLV--IILLC-SVPAVVTIILYGR-----RV 307
 DB 181 SCPEKTPPEPLPPEPGLQMDGFLIITILLCILFALFYLVTLRRNNKKGLKKD 240
 QY 308 AAPARDKSNVDPKKGTSLSDKLSFSTHTLLGQFFQMGTVASWPLTIIIVLSVIVVALA 367
 DB 241 QKSHVTVQVISPSEVTCADKNSQLAQAALSSFOQWGVMAFTPLTVLLSAILIVVLS 300
 QY 368 AGLVFTELTTPDVELWASPNQARSSEKAFHDQFGFPFTNOVILTAPNRSSTRYDSLL 427
 DB 301 AGIKNIELTTPDVELWASPNQARSSEKAFHDTHFDFPFTNOILTAPKEGMYDSLIF 360
 QY 428 GKPNSSGILDLLELLELORLRLQVWSPBAQNISLQDIQVAPLPNDTSLYDCCI 487
 DB 361 GQHNSGIIISKDILIELLELOKQIQNIEMWSKDLNFTATLKQVCFAPLNPNSASTDCAV 420
 QY 488 NSLLQYFQNNRTLLLTANQTLMGQTSQYDWMKHFLYCANAPLTFQDGTALALSCMADYG 547
 DB 421 NSLPQYFQNSLENINAKNMTELGVTKEDWMDHLITCNSPLSFQDITDLGMSCHADYG 480
 QY 548 APVFPFLAIGYKKGKDYSEAEALIMTFSLNYPAGDPLAQAQKWEAELEEMRAFQRM 607
 DB 481 APVIFPLAVGYQ-----MMEMEAELFKIYQEVQKNP 513
 QY 608 AGMFQVTFAAESLEDEINRTTREDLPATSYIVILFYISLALSGYSMSRM-VDSKA 666
 DB 514 STNFTLAYAERSLEDEINRTTREDLPIMISIAVIFLYIAVALGFTSMKRLVVDSKP 573
 QY 667 TGLGVAVALGVMAAMGFYSYLGRSSVILQVVPFLVLSGAGANITFLVLEYORLR 726
 DB 574 LVGLGILVASCALVSSLSGFSWGVGPSLIIIVQVPLVLAQDNITIFVLEYORDVR 633
 QY 727 RGEPEPEVHIGRALGVAPSMILCSLEAICPELGAITMPAVRTFALTSGLAVILDFLL 786
 DB 634 RGEKREBEIGRILGVAPSMILCSLEAICPEFGALSTMPAVKSPALVAVALIDFAL 693
 QY 787 QMSAFPALISDSKQEARLDVCCVQOELEPPQGGSLIGFQKXAYAPPLHMTIR 846
 DB 694 QMTAFPALISDORODSNRCBLICQIKVSRKPTPNEGFLLPFMKKYAPALNRYSR 753
 QY 847 GVALLEFLALFGVSLVSMCHISVGLDOELALPKDSYLLDPLFPLNRYPEVGAUVYFVTL 906
 DB 754 IIVMFPIFMLCGSITLFLHVKVGLDOELAMPKDSIMLKTFEVLKYFVGKAVYIVYTR 813
 QY 907 GYNFSEAGMNAICSSAGCNNSFTQKIOYATEFPBQSYLAIIPASSWVDFTDMLTPSSC 966
 DB 814 GYNFYSVSGMNAVCSVGGCDPYSTQKIOYATEYPELQQLDU----- 855
 QY 967 CLYISGPKKDFCEPSTVNSLNCIKCMGIMTGSVPSVBOFKYLPWFLNDRPNITCKP 1026
 DB 856 -----IKMN 860
 QY 1027 GGLAAVSTSVNLTSQGVLASRPMAYHKPLKNSODYTEALBARBELAANITADLRVPT 1086
 DB 861 SPMLIYAKPFG-----ASRPMAYHTPLNSQETGALMKARELHNITAMRQIQGT 912
 QY 1087 DPAFEVFPYT-----ITNVFERYQVLTILEBGLF 1114
 DB 913 DNFVEFPTTYVLSFSLPLSHYSGDTFIHNDVPSLPAPARTVNFYEQVYIVVEGVL 972
 QY 1115 MSLCLVPPFPAVSCLLIGDLRSGLLNLISYVILVDYVGMALMDISYNAVSLINLVSA 1174
 DB 973 IISLCLPFPVCCLLIGDLRSGLLNLITIMIITVDYVGMALMDIDYNAVALINLVSA 1032

QY 1175 VGMSEVFSHTIRSFALSTKPTLTERAKBATISMGSAVPAVAMTNLPGIILVGLAKQL 1234
 DB 1033 VGISVEFVSHMTRSPALSIKPTHERAKAETQMGSAVFAVAVAMTNLPGIILVGLAKQL 1092
 QY 1235 IOIFFRNLTLITLGLHGLVFLPVILISYVG 1266
 DB 1093 IOIFFRNLTLITLGLHGLVFLPVILISYVG 1124
 RESULT 5
 Q9ULG3_CRIGR
 ID Q9ULG3_CRIGR PRELIMINARY; PRT; 1277 AA.
 AC Q9ULG3;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Niemann-Pick type C1 protein.
 GN Name=NPCL;
 OS Cricetulus griseus (Chinese hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muridae; Cricetinae; Cricetulus.
 OX NCBI_TaxID=10029;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Ovary;
 RX MEDLINE=20127878; PubMed=10660558; DOI=10.1074/jbc.275.6.4013;
 RA Cruz J.C., Sugli S., Yu C., Chang T.-Y.,
 RT "Role of Niemann-Pick type C1 protein in intracellular trafficking of
 RT low density lipoprotein-derived cholesterol.";
 RL J. Biol. Chem. 275:4013-4021(2000).
 DR EMBL; AF182744; AAF31692.1; -; mRNA.
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0005764; C:lysosome; IEA.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0008158; P:hedgenog receptor activity; IEA.
 DR GO; GO:0030301; P:cholesterol transport; IEA.
 DR InterPro; IPR004765; NP_C_type-
 DR InterPro; IPR003392; Patched.
 DR InterPro; IPR000731; SPD_5TM.
 DR Pfam; PF02460; Patched; I.
 DR Trifam; TIGR00917; 2A060601; 1.
 DR PROSITE; PS50156; SSD; 1.
 SQ SEQUENCE 1277 AA; 142817 MW; 6B3D638094B4E156 CRC64;
 Query Match 34.9%; Score 2413; DB 2; Length 1277;
 Best Local Similarity 39.8%; Pred. No. 1e-161;
 Matches 521; Conservative 245; Mismatches 438; Indels 104; Gaps 26;
 QY 11 LVALLLRLAQSEPTTIHQPGYCAFYDEGKRPILSGIMLTSNVCSLSNTPARKITGDH 70
 DB 8 LGLILLILCPAIVFSQ-----SCWYIGEC-----GIAFGKXKNCKSGSPKPLPKDG 55
 QY 71 LILLOKICRLATYGNTOACSGAKQLVLSLEASTTKALLTRCPACSDNFVNLHCNHTCS 130
 DB 56 NDLLQELCRGFFFG-NVSLICDVQQLQTLKSNLQIPMGELSRCPSCFYMLTLFCELTC 114
 QY 131 PMSQPLFINTR-----VAQLGAGQLPAVVAVEAFYGHSPAEQSYDSCSRVRVNAATLAVG 186
 DB 115 PHOSQFLNVTATEDYVDPKTQENKTNVLELYIQSFENEMYNACRDVEAPASNEKALG 174
 QY 187 TMCQVYGSALCNAORMLNFQDGTGNGLAPLDTIT-----FHLPEQAVSGGIQPLNEGVA 241
 DB 175 ILCGDADA-SNAITWIEFMKNNQAPFTTITPISDLPII-----GMEPRNARK 225
 QY 242 RNESEGQDVATCSQDCDAASC-----PALARPQALDSTFYLGQMGPSLVLIILC 292
 DB 226 GCNSESVDVETGRCSCQDCSIVCGPKPQPPPTVPWRILGLDAMY-----VIMWV 274
 QY 293 SVPAVVTIILVGFRAAPARDKSKMVDPKKGTSLSDKLSSTT----- 335
 DB 275 TYMAFLFIFGGLLAVWCHRRRYFV--SEYTPIDSNIAFSSNTSDKGAASCCDPLGAAPD 332

QY 336 -LLGPFQMGKTMWASWPLTILVSVIVVLAAGLVTELTTPDVELMSANSQARSRK 394
DB 333 DCLRMFTKMGAFVYRANPTCVIFFSLVFTACSSGIVFVRYVTNTPVELMSASRSRREK 392
QY 395 AFHDQHGFPFRFNQVILTPAPNRSRYRSDLLGPK-NFSGILDLDLELEBELGRLRH 453
DB 393 EYFDHGFGRFMEQLIQAEPNTSEHITYPPSGSVSGPPLNKEIHLQVLDLQAIAT- 451
QY 454 LQWSPSEARNTSLQDICYAPLPNDTSLYDCCINSLLOYFONNRTLLLTANQTLMGQT 513
DB 452 -SITTSYNNKVTLODICYAPLSPYNNK---NCTIISLVNYPONSHSVL-----DHQVGD 502
QY 514 SQV--DMKHLYCANAPLTFPDGRTALASCMADYAPRPPLLAGYGYKGYSEAEAL 571
DB 503 FYVYADYHTHFLYCVAPAPTSINDTSLDHPCLGTFGGPVPFWLVGGYDQYNNATYLV 562
QY 572 MTFSLNNYPADGPRILAQAALMEAEFLBEMAFORRAGVFOVTFPARESLDEIRRTAE 631
DB 563 ITPFVSNNYNDREKIQRAQAMEKEFTDPYSYKN---PMLTISFLAERSIDELNRESN 619
QY 632 DLPFATSYIVIFLYISLALGSYSWSRVWVSKATLGLGVAVVLGAVMAAGFFSYLG 691
DB 620 DVFTTASVYIMFLYISLALGHSKCSRLVDSKISLAGIILIVLSVACSLGVFSYWG 679
QY 632 IRSSVLIVQVPELVUSGADNIFIVLEYORLPRRGGPREVNHGRALGRVAPSMILCS 751
DB 660 MELTIVIEVIFPLVLAAGVDNIFILVQYQDERQESTLQOQGRILGEVAPFPMFLSS 739
QY 752 LSEALCFGLATLPMPAVRTFALTSGLAYILDLQMSAFVALLSDSKROBASRLDYVC 811
DB 740 PSETAFPFPGALSSMPAVHTFSLPAGLANVLDLQITCFVSLGLGDIRQKRNLDIIC 799
QY 812 CYKPEQLPBPBGQ---EGLLLGFPQKAYAPFLHMTTRGVVLLFLALFGVLSYSMCHI 867
DB 800 CVGQGD---NGRGIQAASESYLFRFPFNKSPAPFLKMDLRPIYIAVFGVLSFIAMNKKY 856
QY 868 SVGLQDQELAPKDSYILDYFLNRFEBGAVRYPTTIGVNFSSBAGNNAICSSAGCN 927
DB 857 EIGLQDSLSMPDSYVIDYFKSLGQVLSGPPVYFLBEGHDYTHKGQNMVCGMGCDN 916
QY 928 FSTFOKIQYATEFPBESYVLAIPASSVWDDFIMLTP--SSCCRLYISGPKDKFCSTVNS 986
DB 917 DSLVQOIFPAALDNTTRIGFAPSSVIDYFPWVAPQSSCCILYNA---THQPCNASTVD 973
QY 987 INCLAKNSIT--MGSVRSVSEQFHKYLPWFLNDRPNIKCPKGLAAYSTSVNLTSDG-QY 1044
DB 974 PTCIR-CRPLTEPGKQRFQCKEFMKFLPMFLSDNPNPKCKGGAAYSAVNIIGDTRY 1032
QY 1045 LASRPMAYHKPLKNSQDYTEALRAARELANITADLRKYPGDDPAEVEVPYITTVFQY 1104
DB 1033 GATYEMTYHTVTKTSADYIDAKKQQLVARNITETWNS-KGSN--YRVPSPVFFVYFQY 1089
QY 1105 YTLIPBEGFLMLSLCLVPTFAVSCLLGLDLSGGLNLNLISVILVDTQFGPALMDISYN 1164
DB 1090 YLTITDITFNLVSLSGISFLVTLVVLGCELSAVIMCTIMILVMGVMWMDISLN 1149
QY 1165 AVSLINLVASGVSEVESHITRSPAISITKPTLWLEBAKEATISMGSAVAGVAMTNLPQI 1224
DB 1150 AVSLVNLVMSGSIYVEFCSHITRAFTMSTKGSRVSRABALAHMGSSVPSGILTFLFGGI 1209
QY 1225 LVTLGLAKQQLIQIFPRRLNLLTLGLGLGVFLPILYISTYGPDVNPA 1272
DB 1210 VYLAFAKQIFEIFYFRMYLAILVLGATHGLIFLPVLLSYGSPVKA 1257

RESULT 6
ID_NPCL_PIG STANDARD; PRT; 1277 AA.
AC P56941;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Niemann-Pick C1 protein precursor.

GN Name=NPCL;
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;
OX NCBI_TaxId=9823;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=21655028; PubMed=11796528; DOI=10.1210/en.143.2.708;
RA Gevry N., Lacroix D., Song J.H., Pescador N., Doblas M., Murphy B.D.;
RT "Porcine Niemann Pick-C1 protein is expressed in steroidogenic tissues
and modulated by camp-";
RL Endocrinology 143:708-716(2002).
CC -!- FUNCTION: Involved in the intracellular trafficking of
CC cholesterol. May play a role in vesicular trafficking in glia, a
CC process that may be crucial for maintaining the structural and
CC functional integrity of nerve terminals (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Found in late
CC endosomes and lysosomes (By similarity).
CC -!- DOMAIN: A cysteine-rich N-terminal domain and a C-terminal domain
CC containing a di-leucine motif necessary for lysosomal targeting
CC are critical for mobilization of cholesterol from lysosomes (By
CC similarity).
CC -!- SIMILARITY: Belongs to the patched family.
CC -!- SIMILARITY: Contains 1 SSD (sterol-sensing) domain.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC EMBL, AF169635; AAD47090.1; -; mRNA.
CC InterPro: IPR004765; NP_C_type.
CC InterPro: IPR003392; Patched.
CC InterPro: IPR000731; SSD 5TM.
CC Pfam: PF02460; Patched; I.
CC TRFAMS; TIGR00917; 2A060601, 1.
CC PROSITE; P550156; SSD; 1.
KW Glycoprotein; Lysosome; Signal; Transmembrane.
FT SIGNAL 1 22
FT CHAIN 23 1277
FT TRANSMEM 271 321
FT TRANSMEM 352 372
FT TRANSMEM 623 643
FT TRANSMEM 655 675
FT TRANSMEM 685 705
FT TRANSMEM 761 781
FT TRANSMEM 834 854
FT TRANSMEM 1099 1119
FT TRANSMEM 1125 1145
FT TRANSMEM 1196 1216
FT TRANSMEM 1228 1248
FT DOMAIN 620 785
FT MOTIF 1274 1277
FT COMBINS 249 259
FT CARBOHYD 70 70
FT CARBOHYD 122 122
FT CARBOHYD 135 135
FT CARBOHYD 185 185
FT CARBOHYD 222 222
FT CARBOHYD 415 415
FT CARBOHYD 452 452
FT CARBOHYD 459 459
FT CARBOHYD 478 478
SQ SEQUENCE 1277 AA; 141963 MW; 2C80D300889F02EB CRC64;
Query Match 34.9%; Score 2412; DB 1; Length 1277;
Best Local Similarity 39.8%; Pred. No. 1.2e-161;
Matches 524; Conservative 234; Mismatches 450; Indels 110; Gaps 25;
QY 7 RGMWLMALLLALQSEPTTTHQPCYCAFYDECGNPELSSWMLTSLSVSCLSNTPARKI 66

Db 4 RGPALGILLILLCPPVQFSG-----SCWTGECG---IASGD-----KRYNCKRYSGPPKPL 51
 QY 67 TGDHLLILCKICRPLTYGPNTOACCSAKOLVSLBASLITKALLTRCPACSDNFVNLHCH 126
 Db 52 PEDGYDLVBLGCGFFFG--NVSLCCDVQQLRTLKNDLQLPLQLFSLRCPCFYNLMLFCB 110
 QY 127 NTCSPNQSILFINTR-----VAOLGAGQLPAAVVAEAFYQHSFPAESYDSCSRVRVPAAT 182
 Db 111 LTCSPNQSQFLNLTATIEDVYDPTVNTQTKTNVKELEFYVGETFANAMYNMCRDVEAPSSNE 170
 QY 183 LAVGTGCGYYSALCNAQRMLNFQGDPTGNGLAPLDIT--FHLEPGQAVSGGIOPINEGV 240
 Db 171 KALGLLGGDAQA--CNATWIEIYMFKNKGQAPFTTPIFSDLPT-----HGMENNNAT 224
 QY 241 ARCNEBQGDVATCSQDCDCAASC-----PALARPOLDSTFYLGQWPGSLVLIIL 291
 Db 225 KCGDEBVDVETGCGSCQDCSIVCGPKRPQPPPPVPRILIGLDAMYIMSSWYAFILVVF 284
 QY 292 CSYFAVVTLLVGFRAVAPARDKSKMYDPKKGTSLSDKLSFSTHTL----- 336
 Db 285 GAFFAVWCY-----RKRYFVSEYTPIDGNIAFVSVNSDKQAFCDDPLGAAP 331
 QY 337 ---LGFQCGWGTWVSWPFLTIVLSVIPVYALAGLVTETLTPDVEILMSAPNSQARAE 393
 Db 332 ERLRLFAQWGAFCVRHPCCVFFSLAFIVACSSGLVFRITTDVDDLMSAFSGQARRB 391
 QY 394 KAFHDQFGPPFPFTNQVILTPARSSRYRDSLILGPK--NPSGLIDLDLLELLELQERLR 452
 Db 392 KEYFDTHGFPFPMEDILIRATNOSHITHPRYAGADVPPGPPSLDILHQVDLQTAIE 451
 QY 453 HLOVMSPEAQRNISLQDICVAPILNPNTSLYDCCINSLQYFONNRTLLLTANQTLMGQ 512
 Db 452 N--ITASYNNEVTTLQDICIAPLSPYNK--NCTILSVLNYFQNSHSLV----- 495
 QY 513 TSGY-----DMKHFLYCANAPLTFQDGTLLALSQADYCAPPEPFLATISGYKXGY 564
 Db 496 DRHQVGFPPVYADYHHPFLYCVRAAPASINDASLHDPCIGTGFGVPFPMVLVCGYDDQNT 555
 QY 565 SEAEALIMTFSLNNYPADGPRLQAOLKMEAFLEEMKAFORBAAGFQVFTFERSLEBE 624
 Db 556 NATALVITFPVNNYNDTEKLOQAQMBEFPINPKANYK--PULTISFMAERSIEDE 612
 QY 625 INRTTABDLPIFATSYIVIFLYISLALGYSYSSMSRWVMSKATLGLGVAVALVGAAMA 664
 Db 613 LNRSNSDLFTTILISYAIMPLYSIALGHKSCSRLLDVSKISLIGILIVLSVASCIL 672
 QY 685 GPFSTYIGISSLVILQVPPVLVSQADNIFIVLEYQRLPRPFGPREVHIGRALGRVA 744
 Db 673 GIFSYIGVPLTLVIVIPPLVAVGVNDIFILVQTYQDRERLQGBETLDOQLGRVIGEVA 732
 QY 745 PSMILGSLSEALICFPIGALTMPMPAVRTPALTSGLAVILDPILQMSAFVALLSIDSKROA 804
 Db 733 PSMFLSSFBETVAFIAGJUSVVPRAVHTFSLFAGMAVLIDBLQITCFVSLISGLIDIRQEK 792
 QY 805 SRLDVCVCYKPOBLPPPGQG---EGLLIGFFQKAVAPFLMHTITRGVULLFLALFGVS 860
 Db 793 NRDVAVCCVQGA---DGAQVQASESCLPFFFGNSVAPLLKQMMRPYIAVAVGLSTS 849
 QY 861 LYSMCHISVGLDQELALPKDSYLLDYFLFLNRYFENGAPRYFTTTLGYNSSGAGNNAIC 920
 Db 850 IAVLNKYEIGLDQSLMPDPSYMDYFQSLSRYLHAGPPVYFVVEEGHNTSLKQGNMVC 909
 QY 921 SASGNNFSFTOKIQVATEPPEQSYLAPASSVNDPFIWLTP--SSCCRLYISGRPKDKF 979
 Db 910 GGLGNCNDSLVQOIFPAQLDNTYTRIGFAPSSWIDYFDWIKFQSSCCRYSN---TDQF 966
 QY 980 CSBTSLNLCNKQMSITM--GSVRPSVEQFHXYLPFLNDRPNIKCPKGGLLAAYSTSVNL 1038
 Db 967 CNASVVDPCIR--CRPLTSEGQRPGQEDPMRFLPMLSDNPNPKGKGSHAAVYSAAVIN 1025
 QY 1039 TSSGQ--VLASRFPAHYKPLKNSQDYTBALPAARELANITADLRKYPGDPDAEVPYPI 1097
 Db 1026 LGNGSGVGIATYFMTYHTVLQASADFTDAMQKARLIASNIT--RTWGLEASSYRVFPPYSV 1082

QY 1098 TNVFFHQYVTLIPBGLFMLSCLVPTFAVSCILLGLDERSGLNLSIYMLIVDTYGFMA 1157
 Db 1083 FYVFEQYVTLVDDITFNLGVLGAIFLVTVLMGCELAATYIMCTIAMILLVNEGVMM 1142
 QY 1158 LMDISYNAVSLINIVASGVSEFVSHITRSPALSTKPMVBRAXATISMSAAVAGVA 1217
 Db 1143 LMGISLNAVSLVNLVMSCGISVEFCHITRAFTLSTKSGRVDPAEALAHMSSVPSGIT 1202
 QY 1218 MTNLPGLIVLGLAKQQLIQIFPFRNLTLITLLGLHGLVFLPVILSYGVPDVPALAL 1275
 Db 1203 LTKFGSIVVLAFAKSGIPIQIFFRMYLAVLIGATHTGLFLFVLVLSIYPSINKASL 1260
 RESULT 7
 Q8MKD8_FELCA PRELIMINARY; PRT; 1276 AA.
 ID Q8MKD8_FELCA
 AC Q8MKD8_08MKD8_01-OCT-2002 (Tremblrel. 22, Created)
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
 DE 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Niemann-Pick C1.
 OS Feline sivestris catus (Cat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Felidae;
 OC Felis.
 OC NCBI_TaxId=9685;
 RN (1)
 RP NUCLEOTIDE SEQUENCE.
 RX MEDLINE=22695580; PubMed=12809639; DOI=10.1016/S1096-7192(03)00074-X;
 RA Somers K.L., Royals M.A., Carstea E.D., Rafi M.A., Wenger D.A.,
 RA Threlk M.A.;
 RL "Mutation analysis of feline Niemann-Pick C1 diseasee.";
 RM Mol. Genet. Metab. 79:99-103(2003).
 DR EMBL; AF03634; BAM27451.1; -, mRNA.
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0005764; C:lysosome; IEA.
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0008158; F:hepatoheg receptor activity; IEA.
 DR GO; GO:0030301; P:cholesterol transport; IEA.
 DR InterPro; IPR004765; NP C type.
 DR InterPro; IPR003392; Patched.
 DR InterPro; IPR000731; SSD_5TM.
 DR Pfam; PF02460; Patched; 1.
 DR TIGRPFAM; TIGR00917; 2A060601; 1.
 DR PROSITE; PS50156; SSD; 1.
 SO SEQUENCE 1276 AA; 141710 MM; C2DD3496472B71EF CRC64;
 Query Match 34.8%; Score 2405; DB 2; Length 1276;
 Best Local Similarity 40.7%; Pred. No. 3,8e-161;
 Matches 535; Conservative 232; Mismatches 461; Indels 86; Gaps 29;
 QY 7 RGLMLALLRLAQSPPYTHQPGYCAFYDECGKRPBLSGLMNTSNVSCLSNTPARKI 66
 Db 4 RGPALGILLILLC---PAQVLAQS--CIWYGBCG--IASGD-----KRYNCKRYSGPPKPL 51
 QY 67 TGDHLLILCKICRPLTYGPNTOACCSAKOLVSLBASLITKALLTRCPACSDNFVNLHCH 126
 Db 52 PEDGYDLVBLGCGFFFG--NVSLCCDVQQLRTLKNDLQLPLQLFSLRCPCFYNLMLFCB 110
 QY 127 NTCSPNQSILFINTR-----VAOLGAGQLPAAVVAEAFYQHSFPAESYDSCSRVRVPAAT 182
 Db 111 LTCSPNQSQFLNLTATIEDVYDPTVNTQTKTNVKELEFYVGETFANAMYNMCRDVEAPSSNE 170
 QY 183 LAVGTGCGYYSALCNAQRMLNFQGDPTGNGLAPLDIT--FHLEPGQAVSGGIOPINEGV 240
 Db 171 KALGLLGGDAQA--CNATWIEIYMFKNKGQAPFTTPIFSDLPT-----HGMENNNAT 224
 QY 241 ARCNEBQGDVATCSQDCDCAASC-----PALARPOLDSTFYLGQWPGSLVLIIL 291
 Db 225 KCGDEBVDVETGCGSCQDCSIVCGPKRPQPPPPVPRILIGLDAMYIMYITMAFLVVF 284
 QY 292 CSYFAVVTLLVGFRAVAPARDKSKMYDPKKGTSLSDKLSFSTHTL 339

Db 285 GAFPLMCKRYKRYFVSEYTPIDSNIAFVSANADRG-----ASCCDLGAAGFEGCLRR 337
Qy 340 PFGQGTWVASWPLTLVLVSVIPVVALAAGLVFTETLTPDVEBLSAPNSQARSEKAHQ 399
Db 338 LFSQMGSCFVRNPGPIIFFSLSLAFCSSGGLVVRVTTPVDLMSAPSSQARLEKEKYPPT 397
Qy 400 HFGPFRINQVILTAENRSSYRYSLLGPK-NFSGILDLDLLELLELOERLRLQVMS 458
Db 398 HFGPFRTEQLIILQPHTSAAHTYQPYPSGSDVFPGLPLALILHQVLDLQTAIEN--ITA 455
Qy 459 PBAQGNISLQDIQVAPLNDPNTSLVQCCINSLLQYQNNRTLLLTANQTLMGQTSQV-- 516
Db 456 SYNNETVTLODLCVAPLSPYNK--NCTLLSVLNTYQNSHSMU----DHEIGDDEFVYA 507
Qy 517 DKMDHFLYCANAPLTFKQDGTALALSCMADYGAVPFPFLAIGKGYKGDYSEABALMTFSL 576
Db 508 DHTHLILYCVARASLNDTSLHDPCLGTFGGVFVFWMLVGGIDDQNNATRLVLTTFYV 567
Qy 577 NNYPAGDPRLAQAQKWEAFLBEMRAFORRMAGMFQVTFABRSLEDEINRTTAEDLPTF 636
Db 568 NNYYNTERLQKAHVMEKEFINFKVKNYKN--PNLTISFTTERSIDEINRESNGDIFVY 624
Qy 637 ARSYVIEFLYISALGYSYSMSRWMDSKATLGLGSAVAVVLGAVMAMGFPSTLGRSSL 696
Db 625 IISYALMFLYISALGHKSCSLRLVDYSKISLIGAILLVLSSVACSLGIFSVYGIPLTL 684
Qy 697 VILQVVPFLVLSGADNIFLTVLEYQRLPRRPGEPREVAHIGRLAGVAPSMILCSSEAI 756
Db 685 IYIEVIFPLVAVGVNIFILVQTYORDERHGETIDQOLGRVLGSAVASMFLSSSEAV 744
Qy 757 CFFLGLATMPAVRTFALTSGLAVILDFLLQMSAPVALLSLDSKQDEASRLDVCCKVQ 816
Db 745 AEFPLGALSKPAVHTSLFAGMAVLIDFLQITCFVSLGLDIXQEKRLDLVLCVRRGS 804
Qy 817 ELDPREQ-GEGLLGFQKAYAPFLHMTTRGVLLFLALRVSLSXMSCHISVGLDOL 875
Db 805 EGGTSVQASBSCFLRFKHSYSPLLKDMWRPVLIAIFGVLSFSAVLNKEIGLDSOL 864
Qy 876 ALPKDSYLLDFLFLNRFEVGAPVFTTLGYNFSSEAGMAAICSSAGCNFSFQKIQ 935
Db 865 SMPDSDSYWMDYFKSL-KYLHAGRPYFVLEBGHDYSLKGQNNVCGMGKCNDDSLVQQLF 923
Qy 936 YATEPEQSYLAIIPASSWVDFIDMLTP-SSCCRLYISGPNKDKFCPSFVNSLNCIKNCM 994
Db 924 NAAQLDSYTRIGRAPSSWIDYFDWKPOSSCCRYSN--TDRFCNASVDPACIR-CR 979
Qy 995 SITM-GSVRPSVEQFKIYLPWLNDRPNIKCPKSGILAAISTSYN-LTSGQVLAASFMAI 1052
Db 980 PLTOEGKORPOGGDFWRFLPMFLSDNPNPKCGKGHAAYSAAVNILGNDTGAVATYFMTY 1039
Qy 1053 HKPLKNSOYTEALRAARELANITADLRKVPGTDA-FEVPFYITTNFYEGYITLDE 1111
Db 1040 HTVLQISADFTDMARKANLANSIT---KTMGLEBSNTRVFPYSYFYFYEGYITLIDD 1095
Qy 1112 GLEMELSLCVPTFAVASCLLGLDLRSGLNLSTIWMILVTVGFMALMDISYNAVSLNL 1171
Db 1096 TIFNLISVLSGAIPLVLVILLGCDLMSAVINCITIAMILVNMFGVMMLMGISLNAVSLNL 1155
Qy 1172 VSAVMSVFEVSHITTSFPAISTKPTWLEBAKXENTIMGSAVPRAGVAMTLPGLVLVGLAK 1231
Db 1156 VMSCGISVEFCSHITTAFTVSMKGRAPQABEALAMGSSVFGITLTLRFKGGLVLAFAK 1215
Qy 1232 AOLIOFFPRNLNLTLGLLHGLVPLVILSYVGVDPNPALAL---EOKRAE 1282
Db 1216 SQIFQIFYFRMYLAMVLLGATHGLIFLPVLVLSITGFSINKAKSLAQEOYKGT 1269

RESULT 8
O9GK52 CANFA
ID O9GK52 CANFA PRELIMINARY; PRT; 1276 AA.
AC O9GK52;
DT 01-MAR-2001 (Tremblrel. 16, Created)

DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Nleam-n-Pick type C1 disease protein.
GN Name=NPCL1;
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
OC Canis.
OX NCB1_TaxID=9615;
RN (1)
RP NUCLEOTIDE SEQUENCE.
RA Muzakawa M., Freeman M.W.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF315034; AAC40873.1; -, mRNA.
DR Ensembl; ENSCAFG0000018183; Canis familiaris.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0005764; C:lysosome; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008158; F:neutrophil receptor activity; IEA.
DR GO; GO:0030301; P:cholesterol transport; IEA.
DR InterPro; IPR004765; NP_C type.
DR InterPro; IPR003392; Patched.
DR InterPro; IPR000731; SSD 5TM.
DR Pfam; PF02460; Patched; I.
DR TIGRfam; TIGR00917; 2A060601; 1.
DR PROSITE; PS50156; SSD; 1.
SQ SEQUENCE 1276 AA; 141695 MW; 07D1327B5BD5681 CRC64;

Query Match 34.8%; Score 2404.5; DB 2; Length 1276;
Best Local Similarity 40.4%; Pred No. 4.2e-161;
Matches 524; Conservative 233; Mismatches 434; Indels 105; Gaps 28;

Qy 33 CAFYECGKNRPLSSGLMTLTVNSCLSNTPARKITGDHILLOKICPRLYTGPNTOACCS 92
Db 25 CVMYGECC---IASGD---KRYNQYSGPRKPLPKQGYDLMLQELCPGLFF-DNYSVCCD 76

Qy 93 AQOLVLSLNSLSTIKALTLTRCAGSDNPNLHCHTCSBNSGLFTINVR---VAQLAG 148
Db 77 VQQLRTLDKSLQLPLQFLSRCPSCFYNLMLNLFCELTCSBROSQFLNVTETEDYVPTNQ 136

Qy 149 QLPVAVAEAFYQHSFAEQSYDSGRVVRPAATLAVGCMGVGYSALCNAQRMILNFOGD 208
Db 137 TKTNYKELOYVGBEFANAMAYACRDVEAPSSNDALGLLCKEBA-CNAITNMLEYFN 195

Qy 209 TNGNLAFL-----DITFHLPEQAVSGIQLPNEGVARNESQGDVATCSQDCQA 261
Db 196 KONGQAFPTIPIFSDLPAN-----GKKPNMNAITKGCDEVDVETAPACSCQDCSV 245

Qy 262 SC-----PALRPAQLDSTFYLGMPGSLVLIILCSVFAYVTILLVGFVAAPARD 312
Db 246 VCGPRQPPAPAPRIRIIGLDAMVYIMVITYMAFLMFGAFAVWCY----- 293

Qy 313 KSKRMVDPKKGTSLSKLSFSTHT-----LLGQFPG-----WGTWVASWPLT 354
Db 294 -RKRYFVSEYTPIDSNIAFVSNAAGDTGBASCCDALGAAPBGLRLFTQWGSFCIRANPQC 352

Qy 355 ILVLSVIPVALAAGLVFTETLTPDVEBLSAPNSQARSEKAHFHQFGPFRTNQVILTA 414
Db 353 IIFGLAFIAACSSGGLVFRVTTNPVDLMSAGSQARLEKEKFDHAFGFPFTEQLIOA 412

Qy 415 PRRSRYDLSLLGPK-NFSGILDLDLLELLELOERLRLQVMSPEAQRNLSLQDICTA 473
Db 413 PHTSVHTYQPYPSGSDVFPGLPLDIGILHQVLDLQTAIEN--ITAYNNEVTLDDICVA 470

Qy 474 PLNPNTSLYDCCINSLLQYQNNRTLLLTANQTLMGQTSQV--DWKDHFLYCANAPLT 531
Db 471 PLSPYNK--NCTILSVLNTYQNSHSMU----DHKIGDFFVYADYRTHLLYCVARAPS 522

Qy 532 FKDGTAALS CMADYGAVPFPFLAIGYKGYKGDYSEABALIMTSLNNYPAGDPRLAQAQ 591
Db 523 LNDTSLHDPCLGTFGGPVFWMLVGGIDDQNNATRLVLTFFPNVNYNTERLQKRAQA 582

Qy 592 WEBAFLBEMRAFORRMAGMFQVTFABRSLEDEINRTTAEDLPIATISYIVFLYISLAL 651

```

Db 583 WEKEPIFVKNYEN---PNLTISTFTERSIEDENLRNSNDVFTVLISYAVMFLYISIAL 639
Qy 652 GSYSSRWVNDKATLGLGVAIVVIGAVMAAGFEFYLIRRSIVLQVPPFLVYSVGA 711
Db 640 GHKSCRFVLDKISLIGILVLSVWCSCIFIFGIPPLVILEVLPFLVAVGV 699
Qy 712 DNIFIPLYEQRLPRRGEPREVHIGRALGVAPSMILCSLSEALICFGLATPMPAVRT 771
Db 700 DNIFILVQYQBERLOGFTLEBOGLGVLEBVAWSMFLSFSEAVAFPLGALSGMPVHT 759
Qy 772 FALTSGIATLDPLOMSAFVALISLDSKQESRLDYCCCVXPEQLPPQCG---EGL 827
Db 760 FSLPAGAAVVIDELQITCFVSLIGDLTKQEKRLDVLCLTQSE---GCTGIQASBSC 816
Qy 828 LDGPFQKAVAPFLHMTTRGVLLFLALRGVSLYSCHISVGLDQSLAPKXSYLDYF 887
Db 817 LRFPFKNSSYSPFLKDMRPVIAVFGILSFSLAVLNKVEIGLDQSLMPDDSYMDYF 876
Qy 888 LFLARYEYVAPVYFVTLTGYNFSEAGMNAICSSAGCNFSTOKIQVATEPEQSYLA 947
Db 877 KSL-KYIHAGPRPYFVLEBGHDYTSLEGQMVCGCMCNDDSLVQOIFSAAQLDNYTHIG 935
Qy 948 IPASSWVDPIDWLTP-SSCCRLYISGPNKDCPCSTVNSLNCIKNCMSITM-GSVRPSV 1005
Db 936 FAPSSWIDYFDWVKPQSSCCRVYNS---TDQFCNAGSVDPACVR-CRPLTQEGKRPPQG 991
Qy 1006 EQRHKYLPWFLNDRPNKCKRKGGLAASVSVNL-SPGOVLASPMAYHPRKRSQDYTE 1064
Db 992 EDPMRFLPMFLSDPNPKCKGHAAGSAVNLGNDSYGATYFMVHTVLQTSADFTD 1051
Qy 1065 ALBAARELANITADLRKVEGTDPAFEVFTYITNVYEQYLTLPGLFMLSLCLVPTF 1124
Db 1052 AMRKATIIANITKTM-SLKGSH--YRPFYSVYVEYEQYLTITIDTINLSVSLGAIF 1108
Qy 1125 AVSCLLGLDLRSGLLNLISIVMLVDTVGPMALMDISYNAVSLINISVAGMSVEFVSH 1184
Db 1109 LVTLLVLLGCELMASVIMCWTIAMILVMFGVMMLMGISLNAVSLNLMSCGISVERCSH 1168
Qy 1185 ITTSFAISTPFTMLERAKENTISGSAVFAGVAMTNLPGILVGLAQOLQIFPFFLNL 1244
Db 1169 ITTAFVYSAGSRVERAEALSHNGSSVFSGITLTKEGVIVLAFAKSQIFQVFTGMYL 1228
Qy 1245 LITLLGLHGLVPLVLTSTVGPVNPALAL-EOKR 1279
Db 1229 AMVLLGNTGHLIFPLVLTSTIGSINAKSLAQOER 1264

RESULT 9
NPCL_HUMAN
AC 015118; 09P130; STANDARD; PRT; 1278 AA.
DT 30-MAY-2000 (Rel. 39, Created)
DT 10-MAY-2005 (Rel. 47, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Niemann-Pick C1 protein precursor.
GN Name=NPCL;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBT_TaxID=9606;
RA NUCLEOTIDE SEQUENCE, VARIANT ILE-642, AND VARIANTS NPCL.
RX MEDLINE=97362323; PubMed=9211849; DOI=10.1126/science.277.5323.228;
RA Carstea E.D., Morris J.A., Coleman K.G., Loftus S.K., Zhang D.,
RA Cummings C., Gu J., Rosenfeld M.A., Pavan W.J., Krizman D.B.,
RA Nagle J., Polymeropoulos M.H., Scurlay S.L., Ioannou Y.A.,
RA Higgins M.E., Comly M., Cooney A., Brown A., Kanecki C.R.,
RA Blanchette-Wackie E.J., Dwyer N.K., Neufeld E.B., Chang T.-Y.,
RA Liscum L., Straus J.F. III, Ohno K., Zeigler M., Carmi R., Sokol J.,
RA Markle D., O'Neill R.R., van Diggelen O.P., Ellender M.,
RA Patterson M.C., Brady R.O., Vanier M.T., Pentchev P.G., Tagle D.A.;

```

```

RT "Niemann-Pick C1 disease gene: homology to mediators of cholesterol
RT homeostasis.";
RT Science 277:228-231(1997).
RN [2]
RP NUCLEOTIDE SEQUENCE, AND VARIANTS.
RX MEDLINE=99355599; PubMed=10425213; DOI=10.1006/dbrc.1999.1070;
RA Morris J.A., Zhang D., Coleman K.G., Nagle J., Pentchev P.G.,
RA Carstea E.D.;
RT "The genomic organization and polymorphism analysis of the human
RT Niemann-Pick C1 gene.";
RL Biochem. Biophys. Res. Commun. 261:493-498(1999).
RN [3]
RP NUCLEOTIDE SEQUENCE, AND VARIANT ILE-642.
RX MEDLINE=21623216; PubMed=11754101; DOI=10.1002/humu.10016;
RA Bauer P., Knoblich R., Bauer C., Finckh U., Hufen A., Kropp J.,
RA Braun S., Kustermann-Kuhn B., Schmidt D., Harzer K., Rolfs A.;
RT "NPC1: complete genomic sequence, mutation analysis, and
RT characterization of haplotypes.";
RL Hum. Mutat. 19:30-38(2002).
RN [4]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA], AND VARIANT ILE-642.
RT TISSUE=Placenta;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Klausner R.L., Feingold E.A., Grouse L.H., Dege J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Heih F.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang Y., Heih F.,
RA Diatchenko L., Marusha K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bernaldo M.F., Casavant T.L., Schreetz T.E.,
RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Frange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., Mcwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Murry D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywiński M.I., Skaleka U., Smalhe D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [5]
RP CHARACTERIZATION.
RX MEDLINE=99128318; PubMed=9927649; DOI=10.1073/pnas.96.3.805;
RA Watari H., Blanchette-Wackie E.J., Dwyer N.K., Glick J.M., Patel S.,
RA Neufeld E.B., Brady R.O., Pentchev P.G., Straus J.F. III;
RT "Niemann-Pick C1 protein: obligatory roles for N-terminal domains and
RT lysosomal targeting in cholesterol mobilization.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:805-810(1999).
RN [6]
RP VARIANT NPD TRP-992.
RX MEDLINE=98299797; PubMed=9634529;
RA Greer W.L., Riddell D.C., Gillan T.L., Girouard G.S., Sparrow S.M.,
RA Byers D.M., Dobson M.J., Neumann P.E.;
RT "The Nova Scotia (type D) form of Niemann-Pick disease is caused by a
RT G3097-->T transversion in NPC1.";
RL Am. J. Hum. Genet. 63:52-54(1998).
RN [7]
RP VARIANTS NPCL GLN-934; LEU-940; ASN-948; LEU-954; TRP-992; ALA-1007;
RX THR-1061 AND VAL-1213.
RX MEDLINE=99452586; PubMed=10521290;
RA Greer W.L., Dobson M.J., Girouard G.S., Byers D.M., Riddell D.C.,
RA Neumann P.E.;
RT "Mutations in NPCL highlight a conserved NPCL-specific cysteine-rich
RT domain.";
RL Am. J. Hum. Genet. 65:1252-1260(1999).
RN [8]
RP VARIANT NPCL THR-1061.
RX MEDLINE=99452593; PubMed=10521297;
RA Millat G., Marais C., Rafi M.A., Yamamoto T., Morris J.A.,
RA Pentchev P.G., Ohno K., Wenger D.A., Vanier M.T.;

```



```

Db      500 GDDFF---VYADYHTHFLYCVRAPASLNDTSLHDPCLGTFGGPVPFWLGGYDDQNN 556
Qy      566 EBAALIMTFSLANNYPAGDPRLAQAOKLMEBAFLBEMKAFORBMAGMFQVTTAERLEDEI 625
Db      557 NATALVITTFVNNYNDTEKQRAQAMEKEFINFVNKNY---PNLTISTABRSJEDL 613
Qy      626 NRTTADLPLEATSYVIFLYISIALGSSYSWSRWVDSKATLGLGVAVALGAYMAAMG 665
Db      614 NRESDDVFTVNVISYALIMFLYISIALGHMKSCHRLLVDSKYSIAGIILIVLSSVASCIG 673
Qy      686 FFEYLGRSLVILQVNPVELVSVGADNIFILEYORLPRRGEPRRHVIGALRVAP 745
Db      674 VFSYIGPLTLVILEVPLVLAVGVNIFILVQAQRDERLQGETLDQOLGRVLEBVAR 733
Qy      746 SMLCSLSEALICFFGLATPMPAPVTFALTSGLAVIDFLQMSAFVALSLSKQOAS 805
Db      734 SMLSSFSSEVAFFLGLALSMVPAVHTPSLPAGLAVPFDILLQITCFVSLGLDIKQOEN 793
Qy      806 RLDDVCCCVKQDELPPPGQ--GEGLLGFFQKAYAPFLHMTTRGVLLFLALRGVSLYSM 864
Db      794 RLDFPCVGAEDGTSVQASESCLEFRFFKNYSGLPLKDMWRPIVLAIFGVLSFSIAVL 853
Qy      865 CHISVGLDQELAPKDSYLLDYFLFLNRYEVGAPYFVTTLLGVNFSSEAGMNAICSSAG 924
Db      854 NKVDIGIDQSLSMPPDSYVWDYFKSISOYLHAGPPVYFVLEBGHDYTSKQNMVCGMG 913
Qy      925 CNNPSTFOKIQVATEPPEQSYLAIPASSWVDDFIDMLTP--SSCCRLYISGNPKDKCPST 983
Db      914 CNDNLSLVQOIFNAAQDLNDNTRYIGRAPSSWIDYDFDWKQPSQCCRV--DNITDQFCNAS 970
Qy      984 VNSLNCNKMSIT--MGSVPSVBOFHAYLPWFELNDRPNKCPKGLAAVSTSVN--LTS 1040
Db      971 VVDPACR--CRPLTPBSKQPOGGDFMRFLPMFLSDNPNKCGKGHAAVSSAVNILLGH 1029
Qy      1041 DGVYLAIRFPAHYHKLKNSODTYBALPAEELANITADLRKYPGTDPAEVPYITITNV 1100
Db      1030 GTVGVGATYFMTYHTVLTQTSADFLDALKKARLIASNTV--ETMGINGS--AYRVFPYSFYV 1086
Qy      1101 FYBOYVLTILEPGLFMLSCLVPTPAVSCLLGLDLRSGLLNLSTYWIIVDTYGPALMD 1160
Db      1087 FYBOYLTITIDTIFNLGVSIGALFLVTVMLGCELSAVIMCATIAVLMFGVMKMG 1146
Qy      1161 ISYNAVSLINLVAAGMSVEFVSHITRSPAISTKPTLBERAKENTISMGSAVAGVAMTN 1220
Db      1147 ISLNAVSLVNLVMSGVSVEFCSHITRAFTVSMKGSVERAEBELAMGSSVSGITLTK 1206
Qy      1221 LPEGLVGLAKAQLIQIFFRLLNLTLLGLLHGLVFLPYILSYGPDVNP 1272
Db      1207 FGGIVVLPAFASQIFQIFFRMYLAVMLGATHTGLFLPYLLSYIGPSVAKA 1258

RESULT 10
Q59GRL_HUMAN PRELIMINARY; PRT; 1289 AA.
ID 059GRL
AC 059GRL
DT 10-MAY-2005 (Tremblrel. 30, Created)
DT 10-MAY-2005 (Tremblrel. 30, Last sequence update)
DE 10-MAY-2005 (Tremblrel. 30, Last annotation update)
DE Nleumann-Pick disease, type C1 variant (fragment).
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Brain;
RA Totori Y., Toyoda A., Takeda T., Sakaki Y., Tanaka A., Yokoyama S.,
RA Ohara O., Nagase T., Kikuno F.R.;
RT "None Title."
RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
EMBL; AB209048; BAD92285.1; -; -.

```

```

FT NON_TER 1 1
SQ SEQUENCE 1289 AA; 143230 MW; 663DE27C13FD801F CRC64;
Query Match 34.8%; Score 2402; DB 2; Length 1289;
Best Local Similarity 39.7%; Pred. No. 6,3e-161;
Matches 525; Conservative 239; Mismatches 452; Indels 106; Gaps 26;

Qy      2 AENGLRGWLLMA-----LLRLAQSEPYTTHQRCYAFDECG-----KNBELSG 47
Db      3 ASAAALRGHMTARGLALGLLILLCPAQVFSQ-----SCVMYGECIAIAGDKRYNCEYSG 57
Qy      48 SLMTSLNVCLENTAPARKITGDHLLILKICCRLYTGPRTQACCSAKOLVLSIASISIRK 107
Db      58 -----PEKPLPKQGYNDVQLCEGFFEG--NVSICCDVQQLQTLKDNQLDLP 102
Qy      108 ALLTRCPASDNFVNILCHNTCSPNQSLFINYTR-----VAOLGAGOLPAVVAEAFYQHS 163
Db      103 QPLSRCPSCFYLLNLNFCBELTCSPPQSOPLANTHEDYVDYPTNQTKMVKELQYVVGQS 162
Qy      164 FAEQSYDSCSYRVVPAAATLAVGTMCGVYGSALCNAQRMLNFQDGTGNGLAPLDIT---- 219
Db      163 FANAMYNACRDVEAPESNDKALGLCGKDADA--CNATNMIIEVMFNKDNQOAFETIIPVES 221
Qy      220 -PHLEPGQAVSGIOLPMBGVARCNEGQDDVATCSCODCAACPAIARQ----- 270
Db      222 DFPV-----KGMENMNNATKGCDESYDEVTAFCSCDCSIICGPKQPPPPAPWTI 273
Qy      271 -ALDSFYLGMQPGSLVLIILICSVFNAVITLL-----VGRFVAPARDKSKM 316
Db      274 LGIDANVTYIMVITTYNAFLVFGAFPAVWCYKRYIVSYTYTIDSIATASV--NASDKGE- 331
Qy      317 VDPKKTGSLSDKLSFSTHTLLGQFPQGWGTWVWASWPLTLLVLSVLPVVALAAGLVTELT 376
Db      332 -----ASCDDPYSAFEGECRLRLFTRWGSCFVRNPGCVIFPSLVITACSSGLVVRVY 385
Qy      377 TDPEVLMSPNGQASSEKAFHQHGGPFRTNQVILTANRSYRDSILLGPK--NFSGI 435
Db      386 TPEVDLMSAPSSQARLEKEYFDQHFQFPFRITQLIRALTLKHITQYPSGADVDFGPP 445
Qy      436 LDDLLLELLELOERLRHLOVMSPEAQRNLSIODICVAPLNPNTSLYDCCINSLHQYQ 495
Db      446 LDIQLIHQVLDIOAIEN--ITASYNETVITODICLALSPNT---NCTILSVANTYQ 500
Qy      496 NNRLLLLTANQTLMTQTSQVMDKHFLYCANAPLTFKDGTAALASCMDYGAVPFFLA 555
Db      501 NSHSVLDHKKGGDF---VYADYHTHFLYCVRAPASLNDTSLHDPCLGTFGGPVPFWL 557
Qy      556 IGGYKGDYSEBALIMTSLANNYPAGDPRLAQAOKLMEBAFLBEMKAFORBMAGMFQVTF 615
Db      558 LGGYDDQNNNATVALTITPPVNNYNDTEKQRAQAMEKEFINFVNKNY---PNLTISF 614
Qy      616 TARSLEDIRNRTTADLPLEATSYVIFLYISIALGSSYSWSRWVDSKATLGLGVAVAL 675
Db      615 TARSLEDIRNRESDDVFTVNVISYALIMFLYISIALGHMKSCHRLLVDSKYSIAGIIL 674
Qy      676 VLGAVMAAMGFSYLGIRSLVILQVNPVELVSVGADNIFILEYORLPRRGEPRRHVIGAL 735
Db      675 VLSVASCISGVSYIGLPILVILEVPLVLAVGVNIFILVQAQRDERLQGETLDQOLGRV 734
Qy      736 IGRALGRVAPSMILCSLSAICFFGLATPMPAPVTFALTSGLAVIDFLQMSAFVAL 795
Db      735 LGRVIGEVAPSMPLSFSSEVAFFLGLALSMVPAVHTPSLPAGLAVPFDILLQITCFVSL 794
Qy      796 SLDKSKQZASRLDDVCCCVKQDELPPPGQ--GEGLLGFFQKAYAPFLHMTTRGVLLFL 854
Db      795 GLDIKROEKNRLDIFCCVGAEDGTSVQASESCLEFRFFKNYSGLPLKDMWRPIVLAIFV 854
Qy      855 ALFGVSLVEMCHISVGLDQELAPKDSYLLDYFLFLNRYEVGAPYFVTTLLGVNFSSEA 914
Db      855 GVLSFSIAYLNKVDIGDQSLMPPDSYVWDYFKSISOYLHAGPPVYFVLEBGHDYTSK 914
Qy      915 GMAAICSSAGCNNSFTQKIQVATEPPEQSYLAIPASSWVDDFIDMLTP--SSCCRLYISG 973

```


Db 915 GQNMVCGMGCGNNDLVQOIFNNAQJDNTRIGFASSWIDYFDMVKQSSCCRV---D 971
Qy 974 PKDKCEPSTVNSLNLCKNCKMSIT-MGSRVPSVEQPHKYLPMFLNDPNIKCPKGGIAAY 1032
Db 972 NITDGCNNSVDPACVR-CRPLTPGKQRPQGGDMRFLPMLSDNPNPKCKGSHAY 1030
Qy 1033 STSVN--LTSDDGYLASRMAHKKPLKNSODYTEALRAARELANITADLRKVPQDPAF 1090
Db 1031 SSAVNTLLHCHGRVGVATFMTYTTVLTQTSADFDAKKRLIASNT-ETMGINGS--AY 1087
Qy 1091 EYEPPTITNVEYEQYTLTBEGFEMSLCLVFPFAVSCLLGLDLSGLNLTSTMYIV 1150
Db 1088 RVFPFSVFFTEQYLLITIDTIFNLGSLGALFLVTWVLGCELSAVIMCATTMVAV 1147
Qy 1151 DTVGFMAIWDISYNAVSLINLVSAVMSYEFVSHITRSFPAISTKPTWLERAKEATISMG 1210
Db 1148 NMFGVWMLGMSLGNANSLVNLVMSGSIYVEFCSHITRAITVSKSRSVERABEALAHMS 1207
Qy 1211 AVFAGVAMTNLPGILVLGLAKAQLIPIFFRLNLITLGLHGLVFLFVILSYGPDVN 1270
Db 1208 SVFSGITLTKFGIVVLAFAKSQLFIYFRMYLAVMLGATHGILFLEVLVLTYSIGPSVN 1267
Qy 1271 PA 1272
Db 1268 KA 1269

RESULT 11.
Q7TMD4_MOUSE PRELIMINARY: PRT: 1277 AA.
ID Q7TMD4; 01-OCT-2003 (Tremblrel. 25, Created)
DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)
DT 10-MAY-2005 (Tremblrel. 30, Last annotation update)
DE Nienman Pick type C1.
GN Mus musculus (Mouse).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
NC NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strauberg R.L., Feingold E.A., Grose L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stables M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uddin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Lonellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richardson S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.C., Murty D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Mair M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strauberg R.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6; TISSUE=Brain;

RA Strauberg R.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC052437; AAH52437.1; -, mRNA.
DR EMBL; BC054539; AAH54539.1; -, mRNA.
DR GO; GO:006897; P:endocytosis; IMP.
DR InterPro; IPR004765; NP_C type.
DR InterPro; IPR003392; Patched.
DR InterPro; IPR000731; SSD 5TM.
DR Pfam; PF02460; Patched; 1.
DR TIGRFam; TIGR00917; 2A060601; 1.
DR PROSITE; PS50156; SSD; 1.
SQ SEQUENCE 1277 AA; 14282 MW; 3B42230AACB564E CRC64;

Query Match 34.7%; Score 2400; DB 2; Length 1277;
Best Local Similarity 40.4%; Pred. No. 8, 7e-16;
Matches 528; Conservative 233; Mismatches 443; Indels 104; Gaps 27;

Qy 11 LMALLRLAQSPPYTIHQPGVCAFYDECGKRPBLSGLSMTLSNVCLSNTPARKITGDH 70
Db 8 LGLLLLLCPAQVFSQ-----SCWVYGRG---IATGD---KRYNCKYSGPKPLPKDG 55
Qy 71 LILQKICPRLYTGPNTOACCSAKQVLSLEASLITKALTRCPACDNFVLHGNCTCS 130
Db 56 YDLVQELCRGLFF-DNVSICCDIQDLQTLKSLQPLQPLSRCSCFNLMLTFCBLTCS 114
Qy 131 PNOSLFINTVTRVAQAGQLPA---VYAEAFYGHSPAESYDSCSRVVPAAATLAVG 186
Db 115 PQSQFLANTATEDYDFDPKQENKTNKELEYFGQSPANMYNACRDVEAPSSNEKALG 174
Qy 187 TMCVYYSALCNQAQWMLNFGDGTNGLAFLDI-----TFHLLPEQAVSGIQPLNEGVA 241
Db 175 LILCGDAPRA-CNATWIEFMKNDGQAPFTIIPVSDLSII-----GMEPRNATK 225
Qy 242 RCNESQGDVATCSCQDCAASC-----PAIARPOLDSTF-----YLGOMPG 283
Db 226 GCNESVDVYTGSCQDCSIYCGPKQPPPPMPRIRIGLDMYIIMVTVYAFLFVFRG 285
Qy 284 SILVLIILCSYFAVNTILLVGRFVAPARDKSRMDPKGTSLSDKLSFTHTLLGQFPQG 343
Db 286 ALLAWCHRRRYFVEYETPIDSNIFSVNNS-----DKSEASCCDPLGAADFCLRMPTK 341
Qy 344 WGTWASMPPLTILVLSVPVALAAGLVETLTDPLVLSAPNSQAESKAFHQHNGP 403
Db 342 WAFVCRNPFTCIFFSLFLFTVCGSLVFOVDTTNVBLMSAPHQALBEKFFDKHGP 401
Qy 404 FFRTOQVILTAPNRSSRYRDSLILLPK-NFSGILDLDELLEBLEQERLRLQVWSPAQ 462
Db 402 FFRTOQVILTAPNRSHVHIEPPAGADVPFGEPPLKELIHQVLDQ-----IAISITAS 456
Qy 463 RN--ISLDICGAPLNDPNTSLYDCINSLLQYFQNNRTLLLTANQTLMGQTSQV--- 516
Db 457 YNNETVTLDDICVADLSPYNK---NCTIMSVLNTYQNSHAVL-----DSQVGD 502
Qy 517 -----DMKDRLPYCANAPLTPFDGTALLSCADVGAAYVFPPLATGCKGKDYSAEALI 571
Db 503 FYIVADYHHPFLYCPRAASLNDTSLHLGPCGTGTGVPFVPLVAGVDDQYNNATLV 562
Qy 572 MTFSLNTPAGDPRLQAQKLEAEFLERAFORRMAGMFOVTFPAERSLEDEIRRTAE 631
Db 563 IFFPNNTYNDERLQRAAMEKERISFYKYKN---ENLTSFPAERSIDELNRENS 619
Qy 632 DLPFATSYIVIFLISLALSSYSSWSRWVDSKATLGLGVAVVGAVMAAMGFFSYLG 691
Db 620 DVFTVIVSYVMFVLISLALGHIQCSRLVDSKISLIGAILVLSVACISGIFSYWG 679
Qy 692 IRSSVILIQVPEVLVLSGADNIFIFVUEYQRLPRPBPBPBVHIGRLARAPBMLCS 751
Db 680 MPLTLVIEVLPFLAVGVNIFLTVQYQDELTQEDTLDDQGRILGAVPMPFSS 739
Qy 752 LSEACFFGALTTPPAVATFALTGLAVIIDLFLQMSAFVALLSLDSKROGASRLDYVC 811
Db 740 PSETAFFFGLASSPAVHTBSLPAGMAVLIDFLQITCFPSLGLDILKROGKHLIDILC 799

Qy	812	CVKQOELEPPPOG-----EGLLGEFOKRYAPFLMLHITRGVLLFTLGLFEVSLYSNMCHI	867
Db	800	CVRGAD---DQGSNASSEYLFRFFPKYNAFALLLKDMLRPLVAVAFGVLSFSAVYANKV	856
Qy	868	SVGLDQELALPKDSYLLDYFLFLNRYEEVGAPVYFVTTLTGYNFSSBAGMNAICSSAGCN	927
Db	857	DIGLDQSLSMFNDSEVIADYFCKSLQYLLHSGCPVYFVLEEYNNYSARKQNNVCCMGCDN	916
Qy	928	FSFPOKIOYAREFPEQSYLALPSSWVNDFLDMLTP--SSCCRLYISGNNKDKFCPSRTYNS	986
Db	917	DSLVOQIFNAELDTYTRVGPAPSSWIDYFDWVSPQSSCCRLY--NVTHQPCNASYMD	973
Qy	987	LNLCNCMSIT-MGVSVPSEVEQFHXYLWPFLNDPENICCPKGGALAAVSTSVNLSDQVL	1045
Db	974	PTCVR-CRPLRPBEGQRQPKGEPMKFLMFLSDNPNPCKGKGHAAYASANNIVGDDTYI	1032
Qy	1046	-ASRFMAVHKPLKNSODYTEALRAARELANITADLRVPGCTDPAFEVFPYTTINVPFQO	1104
Db	1033	GATYFMTHHTLTKTSADYTDAMKKARLTAASITETMBRS-KQSD--YRVFSPSYVVFPEQ	1089
Qy	1105	YLTLIPGLFMLSCTLCTPFAVSCLLGLDRLSGLLNLTSLVMTLVDVVGFMALMDISYN	1164
Db	1090	YLTIIIDDTIFMLSVSLGSIFLVTLTVLVCCELSMAVINICITLAMILNMFYMLLKGISLN	1149
Qy	1165	AVSLINIVSAVGMSEVFPVSHITRSPFAISTKPTWLERAKEATISMGSAVPAGVAMTNLPGI	1224
Db	1150	AVSLINILVMSCGISVEFCSHITRAFTMTSKGRSVARAEALLAHMGSSVFGITLITKFGGI	1209
Qy	1225	LVYGLAKRQQLQIFFFRNLNLTLTGLLGHVLFPVLLSYVGPVDNPA	1272
Db	1210	VLVFAKQIQEIFEYFRMYLANVLLGATHGLIFPVLVLSYVGPSSNKA	1257

	RESULT	12
ID	Q9N0Q0_FELCA	PRELIMINARY; PRT; 1276 AA.
AC	Q9N0Q0;	
DT	01-OCT-2000	(TREMBLrel. 15, Created)
DT	01-OCT-2000	(TREMBLrel. 15, Last sequence update)
DT	01-OCT-2003	(TREMBLrel. 25, Last annotation update)
DE	Niemann-Pick type CI disease protein.	
GN	Name=NPCL1;	
OS	Felis silvestris catus (Cat).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Butheria; Laurasiatheria; Carnivora; Fissipedia; Felidae;	
CC	Felis.	
OX	NCBI_TaxID=9685;	
RN	[1]	
RP	NUCLEOTIDE SEQUENCE.	
RC	TISSUE=Kidney;	
RA	Murakawa M., Freeman M.W.,	
RL	Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AF258783; AAF72187.1; -, mRNA.	
DR	GO: GO:0016021; C:integral to membrane; IEA.	
DR	GO: GO:0005764; C:lysosome; IEA.	
DR	GO: GO:0016020; C:membrane; IEA.	
DR	GO: GO:0008158; F:hedgehog receptor activity; IEA.	
DR	GO: GO:0030301; P:cholesterol transport; IEA.	
DR	InterPro: IPR004765; NP C type.	
DR	InterPro: IPR003392; Patched.	
DR	InterPro: IPR00731; SSD_5TM.	
DR	Pfam: PF02460; Patched; 1.	
DR	TIGRFAMs: TIGR00917; 2A060601; 1.	
DR	PROSITE: PS0156; SSD; 1.	
SQ	SEQUENCE	1276 AA; 141739 MW; CE9B7A0261691256 CRC64;
Query Match	34.7%;	Score 2399; DB 2; length 1276;
Best Local Similarity	40.6%;	Pred. No. 1e-160;
Matches 534;	Conservative 232;	Mismatches 462; Indels 86; Gaps 29;
QY	7 RGMILALLIRLAQSEPTTTHQPGKAYDCCGKNPGLSSMLTLNSVCSLSNTPAKKI	66
DB	4 RGPALGLILLILLC---PAQVLNQS--CIWYGCGC---IASGG---KRYNCKYSGPVKPL	51

QY	67	GGDHLLILLOKICPELYTGPMTQACCSAKQOVSLEASISITALLTRCPACSDNFVNACH	126
Db	52	PKQGYDVLQELCPGFFFDVNSLCCDVQQLQTLKDNILQLPLQFLSRPSCFYNLVNLFCE	110
QY	127	NTCSPNOSLFNTVTR---VAOLGAGQLPAVVAYEAEPYQHSFBAQSYDSCGRVVPAAAT	182
Db	111	LTCSPROSGFLNVTATEDYVDPVTNQTKTNVKELOYITIGSPFANAAMNACGDAVEAPSSND	170
QY	183	LAVGTCGCVYGSAALCAQRMVLFQDGTNGCLAPLDT--FHLEPGQAVSGIOPLNLEGV	240
Db	171	KALGILGCKDAEA--CNATNMIEYFMSKONGAQAPFTIPISDLPT-----HGMEPNNAAT	224
QY	241	ARCHESGDDVATGSCODCAASC-----PALARPQLDSFTFYLGQMPGSLVLIIL	291
Db	225	KGCSESVDEVTGPSCDCDSIVCGPKQPPPEPPVPMWILIGDANTVIMWTYMAFLVTF	284
QY	292	CSVFA-----VWTLLVGFVFA--PARDSKNVDPKKGTSLSDKLSFTHTLIGQ	339
Db	285	GAFPALMCYRKRYVSEYTPIDSNIAASVANADGE-----ASCCDALGAAPEGCLIR	337
QY	340	PFQMGTVVWASWPLTIIIVLSYIPVALAAGLVFTELTTDPVEIWSAPNSQARSEKAFHDQ	399
Db	338	LFQMGSPCVANPPIILFFSLAIFAACSSGLVFRVTTNPVDLMSAPSSQARLEKEYFDI	397
QY	400	HFGPPFTNQVITLAPNBSSTRYDSLLGPK-NSSGILDDLLELLELQERLHLQWS	458
Db	398	HFGFFFRTEQULIDAPHISATHTYPPSGSDVPFGPPLDIALIHQVLDLQTAIEN--ITA	455
QY	459	PEAGRNISSDIOICAPLNPDTSLYDCCINSLLQYQNNRFTLLLTLANQTLMGQTSQV--	516
Db	456	SYNNETVYLQDICAPLSPYTK--NCTILSVLTFQNSHML-----DHEIGDFFVYA	507
QY	517	DMKDHFLYCANAPLTFKDGITALALSCMADYGAPVFPFLAIGGYKQDYSEAEALIMTFS	576
Db	508	DYHHTILYCVARAPASINDTSLHDPCIGTFGAPFPMVLIGGYDQYNNATATAVITFPV	567
QY	577	NNYPAGDPRLOAKLMEBAFLFEENRAQORRAAGCFQVTFTRERLDEBIRNTTMDLPIF	636
Db	568	NNYVNDERLOKAVHEKEPINFKNYKN--PNLITSFTTERSIEDBLNESGDIPTV	624
QY	637	ATSYIVIFLYTSLALGSSMSRWMDSKATLGLAGVAVVLGAAWMAFGPSYIGISSL	696
Db	625	IISYAINMFLYISIALGHIKSSRLNVDSKISLGIAGLITVLVSKACISGIFSYGIPILT	684
QY	697	VILQVPPFLTVSVGADNIFIVLEYQRLPRRPGEPREHVIHRAIGVARAPSMILCSLEAI	756
Db	685	IVIEYIIPFLAVGVNDIFILVQYQYQDERLHGETLQOQGRVLGVAAPSWFLSSFSAY	744
QY	757	CFPLGALTPMEAVRTFALTSGLAVILDELLOMSAFVALLSDSKROEASRLDVCCKVPQ	816
Db	745	APFLGALSKNPAMVHTPSFLPAGMAVLIDPLDITCFVSLDGLDIKROEKNRDLVLCVYGS	804
QY	817	ELPPPGQ--GEGLLGFFQKAYAPFLMHWITRGVULLFLALFGVSYLSMCHISYGLQEL	875
Db	805	EDGTSVQASBSCLFRLFHGSYSPLLLDKMWRIVIALFVGVLSPSVAVLNTVEYLGLOSL	864
QY	876	ALPKDSYLDVFLFNRYFEVIGAPVYFTTLGYNFSSBAGNATLCSASGANNPFSFKIQ	935
Db	865	SMPDSDYMDYFKSL-KTLHAGPPVYFVLEGGHITSLKQGMVYCGGMCNNDSIVQOIF	923
QY	936	YATFPEOSYLAIPASSWVDDFIDMLTP--SSCCRLYISGPNKDKRCPSTVNSLCLKNM	994
Db	924	NAAGUDSYTRIGFAPSSWIDDYFPMVKQSSCCRHYNS--TDRFCAASVADPAICIR--CR	979
QY	995	SITM--GSYRPSVEQFHKTLPWFLNDRPVIKCPKGGLAAYTSYV--LTSQGVASLRPMAY	1053
Db	980	PLTQEGKRPQCGCFMRFLPMFLSDNPNPKCGKGGHAYSSAVNILGNDTVGATYEMTY	1039
QY	1053	HKPLKNSDYEAALRAARELANNTADLRKVPGTDA--FEVFPYITVVFQYQYLTIPE	1111
Db	1040	HTVLQTSADFTDAKKANLISNIT---KTMGLBGSYRFRFPISVVFYFQYVILTIID	1095

```
Qy 1112 GFPMISLCLVPTFPAVSCLLIGDLBSGLNLSTIWMIVDTYGFPMALNDISTYNAVSLNL 1171
Db 1096 TIFNLISVSIKGAIFLVYLVLLIGCDLMSAVIMCITIAIMILVNFQVWMLWGISINAVSLNL 1155
Qy 1172 VSAVGSVEFVSHITSPASTKPTWLEBAKATISMSGSAVAGVAMTNLPGLIVGLAK 1231
Db 1156 VMSCGISVFCSCHITRAFTVSMKGSRAQAEALAMSGSVFSRGITLTTFGGIVLAAPAK 1215
Qy 1232 AQLIQIFFRNLNLITLLGLHGLVLPVILSYVGDVNPALAL--EQKRAEE 1282
Db 1216 SQIFQIFFRMYLANVLLGATHGLIFLPVLLSYIGSINKAKSLATQEQYKTE 1269

RESULT 13
Q8MI49_FELCA
Q8MI49_FELCA PRELIMINARY; PRT; 1276 AA.
AC 08MI49;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Mutant Niemann-Pick C1.
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Felidae;
OC Felis.
OC NCBI_TaxID=9685;
OX (1)
RN NUCLEOTIDE SEQUENCE.
RP MEDLINE:22695580; PubMed:12809639; DOI=10.1016/S1096-7192(03)00074-X;
RA Somers K.L., Royals M.A., Carstea E.D., Rafi M.A., Wenger D.A.,
RA Thirall M.A.;
RT "Mutation analysis of feline Niemann-Pick C1 disease.";
RL Mol. Genet. Metab. 79:99-103(2003).
EMBL: AF503633; AAM27450.1; -; mRNA.
DR GO: GO:0016021; C:integral to membrane; IEA.
DR GO: GO:0005764; C:lysosome; IEA.
DR GO: GO:0016020; C:membrane; IEA.
DR GO: GO:0008158; F:hedeghog receptor activity; IEA.
DR GO: GO:0030301; P:cholesterol transport; IEA.
DR InterPro: IPR004765; NP C type.
DR InterPro: IPR003392; Patched.
DR InterPro: IPR000731; SSD 5TM.
DR Pfam: PF02460; Patched; I.
DR TIGRPFAM: TIGR00917; 2A060601; 1.
DR PROSITE: PS0156; SSD; 1.
SQ SEQUENCE 1276 AA; 141694 MW; CB356221419599BF CRC64;

Query Match 34.7%; Score 2395; DB 2; Length 1276;
Best Local Similarity 40.6%; Pred. No. 2e-160;
Matches 534; Conservative 232; Mismatches 462; Indels 86; Gaps 29;

Qy 7 RGMILMALILRLAQSEPTTTHQPGYCAFYDECGKPELSGIMLTUSNCLSNTPARKI 66
Db 4 RPPALGLLILLC---PAQVLAQS--CIYWGEG---IASGD---KRYNCKSGEPKPL 51
Qy 67 TGDHLILLOKICPRLYTGPNTOACGSAKOLVSIASLITKALLTGPCASDNFVNLCH 126
Db 52 PRQGVDLVQELCPGFPP-DNVSICDVQLOQLKDNLOPLQPLSRCPSCFYLVNLPCB 110
Qy 127 NTCSPNQSLFIVNTR-----VAQLGAGQLPAVVAYEFYQHSFPAEGSYDSCSRVVAAT 182
Db 111 LTCSPNQSLFIVNTR-----VAQLGAGQLPAVVAYEFYQHSFPAEGSYDSCSRVVAAT 170
Qy 183 LAVGTCGVYGSALCAQARMNLNPOGDTGNGLAPLDT--FHLLPQGAAGSGIOPANEV 240
Db 171 KALGLICGDARA-CNATWIEIYMSKDNQABFTTPIPSDLPT-----HGEPNANAT 224
Qy 241 ARCNESQGDVAATCSQDCQASAC-----PAIARPOLDSTFYLGQMPGSLVLIIL 291
Db 225 KCGDSEVDVETGCGQDCSIVCGKPPPPVPRRIIGLDAMYIIMITWAFLLVVF 284
Qy 292 GSVA-----VTTIIVGRVA---PARDKSKVDPKKTGSLSDKLSTSTHTLLQ 339
Db 292 GSVA-----VTTIIVGRVA---PARDKSKVDPKKTGSLSDKLSTSTHTLLQ 339
```

```
Db 285 GAFFMLCKRKRYFVSETPIDSNIAFSVNNDRGE-----ASCQALGAFFGCLR 337
Qy 340 FFGQGTWVAWNPILTVLSVIPVVALAGLVFTLTDVPELWSPNSQASERAFHQ 399
Db 338 LFSQGSFCVNRPGPIIFSLAFIAACSSGLFVAVTNPNVDLWSPASQALREKEYFD 397
Qy 400 HGGPFRTNOVILTPKNSSTXYDSSLGPK-NFGIILDDLLELLELQELRLQWS 458
Db 398 HGGPFRTNOVILTPKNSSTXYDSSLGPK-NFGIILDDLLELLELQELRLQWS 455
Qy 459 PEAQNISLHODICVAPLPNDNTSLVDCCINSILQYFQNNRTLLLTANQTLMSQSY- 516
Db 456 SYNMETVTLQDICVAPLSPYK--NCTIISLVANFQSHSNL-----DHETGDDPFVYA 507
Qy 517 DKQHFLLVCANAPLTFKQGTALALSQMDYGAVPFPAIGYKQDYSEBALIMTSL 576
Db 508 DYHTLLVCVAPASLNDTSLLDHDCLTGFGGPPVPMVLGGYDQNYNNAATLVITFPV 567
Qy 577 NNYPRGDRILAQAKMEBAFLBEMRAFORBAMGMQVFTABRSLEDELNRTAADLPF 636
Db 568 NNYPRGDRILAQAKMEBAFLBEMRAFORBAMGMQVFTABRSLEDELNRTAADLPF 624
Qy 637 ATSYIVIPLYTSLALGSYSMSRWVWDKATLGLGVAVVLGAWMAAGFSGYLGRSL 696
Db 625 IISYAIMFLYISIALGHKSCSRLVDKISLGIGILVLSVACSLGIFSYGILPLT 684
Qy 697 VILQVPPFLVLSVGADNIFIVLEYQRLPRRGPBREVHIGRALGVAPSMLLCSLSEA 756
Db 688 IYIEVIFPLVALAVGDNIFILVQYQRODBRLHGETLDQDLGVGEVAPSMFLSFSSEAV 744
Qy 757 CFFLQALPMPAVRPFALTSGLAVIDLTLQMSAVALLSLDSKQOBSRLDYCCCYKQ 816
Db 745 AFFLALSMAPVAHFPSLPAGMAVIDPLQITTCVSLGLDIDKQEKRRDLVLCVGS 804
Qy 817 ELPRPGQ--GEGLLGFFOKAVAPFLHWTIRGVVLLFALPGVSLGHSCHISVGLDEL 875
Db 805 EGGTSVQASSECLFPLFKHSISPLLLKQMPRIVAIFGVLSFSAVNLKRIEIGDSL 864
Qy 876 ALPKDSYLLDYPLFLNRYFEVGAQVYFVTTIGVNFSSBAGMAICSSAGCNNSFTQKIQ 935
Db 865 SMPDSDSYMDYFKSL-KYLHAGRPVYFVLEBHDYTSLKQGMVCGMGCNNDLSLVQQLF 923
Qy 936 YATPEPQSYLAIPASSWVDPEIDWLT-P--SSCCRLYISGPNDRKCPSTVNSLNLCKM 994
Db 924 NNAQDLSYTRIGFAPASSWIDVFDWVKPQSSSCRYVNS--TDRECNASVNDPACIR-CR 979
Qy 995 SITM-GSVRPSVEQHKLPMFLNDRPNIKCKPGGLAAYSTSN-LTSDQVLAGRFMAX 1052
Db 980 PLTQEGKQRPQGGDMRFLPMFLSDNPCKCGKGAAVSSAVNLGMDTGATFYMTY 1039
Qy 1053 HKPLKNSODYTEBALRAABELANITADLRKVGTDPA--FEVPPYTTIVNFYQVYLTPE 1111
Db 1040 HTVLQTSADPFTAMKMANLIASNIT---KTMGLGSNVRYVPVYFVYQVLYTIID 1095
Qy 1112 GFPMISLCLVPTFPAVSCLLIGDLBSGLNLSTIWMIVDTYGFPMALNDISTYNAVSLNL 1171
Db 1096 TIFNLISVSIKGAIFLVYLVLLIGCDLMSAVIMCITIAIMILVNFQVWMLWGISINAVSLNL 1155
Qy 1172 VSAVGSVEFVSHITSPASTKPTWLEBAKATISMSGSAVAGVAMTNLPGLIVGLAK 1231
Db 1156 VMSCGISVFCSCHITRAFTVSMKGSRAQAEALAMSGSVFSRGITLTTFGGIVLAAPAK 1215
Qy 1232 AQLIQIFFRNLNLITLLGLHGLVLPVILSYVGDVNPALAL--EQKRAEE 1282
Db 1216 SQIFQIFFRMYLANVLLGATHGLIFLPVLLSYIGSINKAKSLATQEQYKTE 1269

RESULT 14.
NPCI MOUSE STANDARD; PRT; 1278 AA.
ID NPCI MOUSE
AC 035604; 035605;
DT 30-MAY-2000 (rel. 39, Created)
DT 30-MAY-2000 (rel. 39, Last sequence update)
```

DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Niemann-Pick C1 protein precursor.
DE Name=Npcl;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Mus.
OC NCBI_Taxid=10090;
[1]
RN NUCLEOTIDE SEQUENCE.
RX MEDLINE=97362324; PubMed=9211850; DOI=10.1126/science.277.5323.232;
RA Loftus S.K., Morris J.A., Carstea E.D., Gu J.Z., Cummings C.,
RA Brown A., Ellison J., Ohno K., Rosenfeld M.A., Tagle D.A.,
RA Pentchev P.G., Pavan W.J.;
RT "Murine model of Niemann-Pick C disease: mutation in a cholesterol
RT homeostasis gene";
RT Science 277:232-235 (1997).
RN [2]
RN SUBCELLULAR LOCATION.
RX MEDLINE=99145606; PubMed=9990080; DOI=10.1073/pnas.96.4.1657;
RA Patel S.C., Suresh S., Kumar U., Hu C.Y., Cooney A.,
RA Blanchette-Mackie E.J., Neufeld E.B., Patel R.C., Brady R.O.,
RA Patel Y.C., Pentchev P.G., Ong W.-Y.;
RT "Localization of Niemann-Pick C1 protein in astrocytes: implications
RT for neuronal degeneration in Niemann-Pick type C disease";
RT Proc. Natl. Acad. Sci. U.S.A. 96:1657-1662 (1999).
CC -!- FUNCTION: Involved in the intracellular trafficking of
CC cholesterol. May play a role in vesicular trafficking in glia, a
CC process that may be crucial for maintaining the structural and
CC functional integrity of nerve terminals.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Found in late
CC endosomes and lysosomes.
CC -!- TISSUE SPECIFICITY: Expressed predominantly in perilymphatic
CC astrocytic glial processes. Also expressed in heart, spleen, lung,
CC liver, skeletal muscle, kidney, testis.
CC -!- INDUCTION: Activated by the drugs progesterone and U-18666A which
CC block cholesterol transport out of lysosomes and by the
CC lysosomotropic agent NH4Ac.
CC -!- DOMAIN: A cysteine-rich N-terminal domain and a C-terminal domain
CC containing a di-leucine motif necessary for lysosomal targeting
CC are critical for mobilization of cholesterol from lysosomes.
CC -!- SIMILARITY: Belongs to the patched family.
CC -!- SIMILARITY: Contains 1 SSD (sterol-sensing) domain.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
DR EMBL: AF003348; AAB63372.1; -; mRNA.
DR EMBL: AF003349; AAB63373.1; -; Genomic DNA.
DR FIR: T30188; T30188.
DR Ensembl: ENSMUSG0000024413; Mus musculus.
DR MGI: MGI:1097712; Npc1.
DR GO: GO:0006897; P: endocytosis; IMP.
DR InterPro: IPR004765; NP_C_type.
DR InterPro: IPR003392; Patched.
DR InterPro: IPR000731; SSD_5TM.
DR Pfam: PF02460; Patched; 1.
DR TIGRfam: TIGR00917; 2A060601; 1.
DR PROSITE: PS50156; SSD; 1.
DR GlycoProtein: Lysosome; Signal; Transmembrane.
KM SIGNL 1
FT CHAIN 23
FT 1 23
FT 24 1278
FT TRANSMEM 271 291
FT TRANSMEM 352 372
FT TRANSMEM 623 643
FT TRANSMEM 655 675
FT TRANSMEM 685 705
FT TRANSMEM 761 781
FT TRANSMEM 834 854
FT TRANSMEM 1099 1119

FT TRANSMEM 1125 1145 Potential.
FT TRANSMEM 1196 1216 Potential.
FT TRANSMEM 1228 1248 Potential.
FT DOMAIN 621 765
FT MOTIF 1275 1278
FT COMPAS 250 260
FT CARBOHYD 71 71
FT CARBOHYD 123 133
FT CARBOHYD 138 138
FT CARBOHYD 186 186
FT CARBOHYD 223 223
FT CARBOHYD 229 229
FT CARBOHYD 415 415
FT CARBOHYD 460 460
FT CARBOHYD 479 479
FT CARBOHYD 525 525
FT CONFLICT 474 474
FT CONFLICT 479 479
SQ SEQUENCE 1278 AA; 142889 MW; 43C71CE47D283674 CRC64;
Query Match 34.5%; Score 2385; DB 1; Length 1278;
Best Local Similarity 40.2%; Pred. No. 1e-159;
Matches 525; Conservative 233; Mismatches 443; Indels 104; Gaps 27;
QY 14 LLLRLAQSEPPYTTIHQPGYCAFYDECGKNPELGSIMTLNVSCLSNTPARKITGDHLIL 73
12 LLLLCPAAVFSSQ-----SCWYGEBCG---IATGD-----KRYNCKYSGPKPLPKDGYDL 59
QY 74 LQKICERLYTGRPTQACCAKQVLSIASLSITKALLTRCPACSDPFWNLHCINTGSPNQ 133
60 VQELCPGLFP-DNVSICCPIQDLQTLKSNLQPLQPLSLRSCFSCFVNLMTLFCGLTSPHQ 118
QY 134 SLPIINTVRAQAGAGOLPA-----VVAYEAFYOHSEFASQSYDCSRVRVPAATLAVGTMC 189
119 SQFLNVTAREDFPDEPTENKTNVKELEYVGOSFANMANTMCRDYAEASSNEKALGLLC 178
QY 190 GYVGSALCAQAKWLNFGDGTGNGLAPLDI-----TFHLLPQAVSGGIQPLNEGVARN 244
179 GNRARA-CNATMWIEFMFKNGQAFTIIPVFSDLSTL-----GMEPMNATKGCN 229
QY 245 ESGGDDVATCSCODCAASC-----PAIARPAQLDSTF-----YLGQMGSLV 286
230 ESDVETVGCSCQDCSIVCGPKRQPPPPMPPRIMGLDMYVIMVTVYAFVLFVFGALL 289
QY 287 LIIILSVFAYVTILLVGRVAPARDKSNVDPKKGTSIDKLSFSTHTLLTGGFFQCGMT 346
290 AWCCHRRRYFSEYTPIDSNIFSVNS-----DKGEASCCDPLGAAPDCLRMFTKMA 345
QY 347 WVASWPLTTLVSVIPVALAGLVFTLTPDVELMSAPNSQAREKAFHDQFGPFR 406
346 FQVRNPCTCIIFSLARITVCGSGLVFVQVTTNPVELMSAPHSQARLEKEVPDGHFGPFR 405
QY 407 TNOVITLAPNRSSRYRSDLSLLGPK-NFSGILDLLELLELQERLRHQVWSPKQRN- 464
406 TEOLIIQAENSTVHYIEPPAGADVPFGPLNKEILHQVLIHQ-----IAIEISITASVNN 460
QY 465 --ISLDICVAPLNPNPTLVYCCINSILYQFONRNTLLLTANQTLMGQTSQV----- 516
461 ETVTLQDVCVAPSPYKN--NCTIWSVLANTRQNSAVL-----DSQVGDFFYI 506
QY 517 --DMKDFPLVCANAPLTFPDGTALASCNADYGAPVFPPLAIGYKXKDYSEAEALIMTF 574
507 YADYHHTFLYCRAPASLNDTSLHGPCLGTFGQPPVPLVGVGGYDDQYNNATALVITF 566
QY 575 SLNNYPADGPRLAQAKLEBAFLPEMRAPQRRMAGFQVTFPAERSLBEINFTTADLP 634
567 PVNNYNDTERLQRAWAMEKEFISFYKNYKN--PWLITISFPAERSIBBELNRESNDVFE 623
QY 635 IPATSYIVFLYISLALSGYSRMYNDSKATLIGGAVVVLGAAMAASFYSYGISS 694
624 TVIISTVWVPLYLISLALGIIQSCSRLVDSKSLGAGLILVSVAGSLGIFSTYGMPL 683
QY 695 SLVILQVVPFLVLSVGADNIFLVLEYQRLPRRPGEPREVIHGRALGRVAPSKMLCSLSB 754

[illegible]

Query Match	34.3%	Score 2367.5	DB 2	Length 1277
Best Local Similarity	40.1%	Pred. No. 1.7e-158		
Matches 517	Conservative 227	Mismatch 445	Indels 101	Gaps 26
QY	33	CAFYDECGKNPELGSIMLTLSNVSCLSNTPARKITGTHLLLOKICPRLTYGPNTQACCS	92	
DB	25	CIMYGECC---IASGD---KRYNCRISGPEPELPQDGYDLVQELCEGFFFG--NVLSCD	76	
QY	93	AKOLVLSIASLITKALLTRCPACSDNFVNLHCNHTCSNQSLFEINYTRAQT--GAGQ	149	
DB	77	VQGHHTLKNDLQPLQPLQLSRCPSCFYVNLVPLGELTCSPROSGELNVTATEDYVDPAHQ	136	
QY	150	LPAVVAEARY-QHSPAEQSYDCSRRVVAATLAVTMCQGYGSLCNAAQRLNQCQD	208	
DB	137	TKTNVKELOQYVESEFANMAMTNSCRDVEASSNEKALGLLCGRBASA-CNATNMIETMFN	195	
QY	209	TGNGLAFLDYLT--FHLEPGQAVSGIGIOPINEGVARNESQGDVATVASCODCAASCPAI	266	
DB	196	KDNQQAEPFITTPVPSPDLPT-----HGMEPMNNAATKGCDESDVEYTGSCSCODCAVCGPK	250	
QY	267	ARPD-----ALDSTFYLGOMPGSLVLIILCSYFAVVTILLVG--FRVAPARDK--	313	
DB	251	PQPEPEVPMKILGLDAMY-----VIMSTYMAFLVFGFAFPAVMCYRKRYF	298	
QY	314	-----SKWVDPKKGTSLSDKLSFSTHTLLGQFPQCGTGVNAEMPLTIVLVS	360	
DB	299	VSEETPIDGNIPEISINASDKGPPCCDPLAAEPAHRLFEWMSGCVHPPGVFFSV	358	
QY	361	IPVVALAGLVFELTDPPELMSAPNSQARSEKAFHDOFGPPFRNQVILTPAPNSSY	420	
DB	359	AFIACSSGLVFIQVTTDPVDLMSAPBSQARLEKEYDTHFGPPFRLEOLITRAPHPRH	418	
QY	421	RYDSLILGPK-NFSGILDDLLELLELQERLRLQWSPDAQNISLQDIQVAPLNDN	479	
DB	419	IYERYPGADVDFGPPLAVNIHQVLDLQVAIE--SITASYNMETVTLRDIQVAPLSPYN	476	
QY	480	TSLYDCCLNSLQYFQNNRLLILLTANQTLMGQTSQV--DMKHFLYCAAPLTFKQCTA	537	
DB	477	---ONCTILSLVNFQNSHSLV-----DHQVGDFFYAADYHHTHFLVWAPASLNTSL	528	
QY	538	LALSCHMDYGAVPYPPFLAIGYKQDYSEABALIMFSLNNVPADGRLQAOLMEBAFL	597	
DB	529	LHDPCLGTGGPAPFPMVLVGGYDDQYNNAITALVITTPVANNYNDRTEKLRQAQMEKEFI	588	
QY	598	ENMRAFORMAQMFQVTFEASLEDEINFTTAEADLPIFATSYIVIFYLSLAYSYSW	657	
DB	589	NFVQNYEN---PNLTISFKAERSIEDLBNESNDVFTVLISGVMEFLYSIALGHKISC	645	
QY	658	SRVWVDSKATLGLGCAVAVLGAVMAAMGPFSSYIGRSLSVILQVNPPLVUSYQADNFI	717	
DB	646	RRLVNDSKILLAGVILVISPACSDGPFSSYMGSSPLTLVIEVIEPLVAVAGDNI	705	
QY	718	VLEQYRLRRRPGEPREYHIGRALGRVAPSMULCSLSBAICFPLGALTPMPAVRTFALTS	777	
DB	706	VQTYQIDBRLOGETLDDQVGRVLTGEVAPSNFLSSFAETVAFPLGGLSVMPRAVHTFSLPAG	765	
QY	778	LAVILDELQMSAFVALLSDSKRQASRLDVCVCVPOELPPPG-----QSGELL	828	
DB	766	MAVLIDELQITGCVSLTGLDRIKQEKNGQDVLCCV-----GGAADAGIQASBSCL	817	
QY	829	LGFQQAIAAPPLHMTIRGVVLLFLALFGVSLYSKMHISVGDQDELAKRDSITLDYFL	888	
DB	818	FRFRNSYAPLLKLDMMRPLVAVVFGVLSFSLAVLNKVEIGLDQSLSPMDSDSVTYTYFQ	877	
QY	889	FLNNYFEVGA.PYFVVTTLTGYNFSEEAQMNALCSAGCANNPSFOCKIQYATPEFQSYLAI	948	
DB	878	SLANDYLHAGPPVYFVLEEGHDYISTKQNNVCGGLGCGNDSLYQDVTTAAQDSYTHIGF	937	
QY	949	PASSWVDFTDMLTP--SSCCRLYISGENKQKFCFSTYNSLNLCKNCSIT--MGSVRPVE	1006	
DB	938	APSSMIDYDPMVYRQSSCCRITNS---TEQFCNASVYVNPCTV-CAPLTPBEGRORPQGA	993	
QY	1007	QFHKLFWFLNDRPNICPKPGGLAAYSTVNLISDG-QVLAISPMAYHKPLKNSQDYTEA	1055	

THIS PAGE BLANK (USPTO)